In AI We Trust? Effects of Agency Locus and Transparency on Uncertainty Reduction in Human–AI Interaction

Bingjie Liu

California State University, Los Angeles, Los Angeles, CA 90032-8300, USA

Artificial intelligence (AI) is increasingly used to make decisions for humans. Unlike traditional AI that is programmed to follow human-made rules, machine-learning AI generates rules from data. These machine-generated rules are often unintelligible to humans. Will users feel more uncertainty about decisions governed by such rules? To what extent does rule transparency reduce uncertainty and increase users' trust? In a 2 × 3 × 2 between-subjects online experiment, 491 participants interacted with a website that was purported to be a decision-making AI system. Three factors of the AI system were manipulated: agency locus (human-made rules vs. machine-learned rules), transparency (no vs. placebic vs. real explanations), and task (detecting fake news vs. assessing personality). Results show that machine-learning AI triggered less social presence, which increased uncertainty and lowered trust. Transparency reduced uncertainty and enhanced trust, but the mechanisms for this effect differed between the two types of AI.

Lay Summary

Machine-learning AI systems are governed by system-generated rules based on their analysis of large databases. These rules are not predetermined by humans. Furthermore, they can sometimes be seen as difficult to interpret by humans. In this research, I ask whether users trust the judgments of such systems that are driven by machine-made rules. The results show that when compared with a traditional system that was programmed to follow human-made rules, machine-learning AI was perceived as less humanlike. This led users to be more uncertain about the decisions produced by the machine-learning AI system. This also decreased their trust in the system and their intention to use it. Transparency of the rationales for its decisions alleviated users' uncertainty and enhanced their trust, provided that the rationales are meaningful and informative.

Keywords: Machine Learning, Agency Locus, Agency Attribution, Transparency, Uncertainty, Trust

doi:10.1093/jcmc/zmab013

Corresponding author: Bingjie Liu; e-mail: bliu3@calstatela.edu

Associate Editor: Mike Yao; Received: 12 November 2020; Revisions received: 25 June 2021; Accepted: 14 July 2021; Published: 9 September 2021

Introduction

Artificial intelligence (AI), defined as machines that can perform tasks that require human intelligence (Turing, 1950), has been widely deployed to make decisions for humans. For example, AI is now used for detecting misinformation, identifying problematic online profiles, personalizing recommendations, and deciding who gets jobs and who is eligible for parole (Araujo, Helberger, Kruikemeier, & De Vreese, 2020). Although these AI systems are designed to make sound decisions, a majority of Americans distrust AI-made decisions (Smith, 2018). Among the factors that lead the public to doubt AI's decisions, one often-cited reason is users' uncertainty about its decision-making processes (Lee, 2018). That is, they lack the knowledge necessary to predict and explain AI-made decisions (Berger, 1987). What characteristics of AI may induce uncertainty and make it untrustworthy? How should AI be designed to serve human goals? The increasing delegation of decision-making to AI has created an urgency to answer these questions.

One feature of modern AI—the use of machine-learning techniques—may particularly induce uncertainty among users. Unlike traditional AI systems that are programmed by humans to follow human-made rules (e.g., expert systems in the 1970s), modern machine-learning AI systems define and modify decision-making rules through learning from data. These machine-generated rules are often unknown and might be fundamentally different from human logic (Mittelstadt, Allo, Taddeo, Wachter, & Floridi, 2016). Research so far suggests that individuals tend to have simulated human intelligence and use their knowledge about humans (i.e., anthropocentric knowledge) to reduce uncertainty about machines; but this process is contingent on the applicability of anthropocentric knowledge to the machines they interact with (Epley, Waytz, & Cacioppo, 2007). Since machine-learning AI systems' decision-making does not directly reflect human knowledge and principles, will this fact inhibit users from simulating human intelligence, thereby increasing their uncertainty and decreasing their trust in machine-learning AI systems? To answer this question, the present study conceptualizes the difference between the two types of AI's decision-making rules (i.e., human-made vs. machine-generated) as a shift of AI's "agency locus" from human to machine and investigates this shift's effects on users' uncertainty about and trust in AI-made decisions.

To mitigate users' uncertainty, researchers in Explainable AI (XAI) have been working toward enhancing AI transparency by building AI systems that can explain their decisions to the users (Miller, Howe, & Sonenberg, 2017). Although some research has found that AI transparency could increase users' satisfaction (Shin & Park, 2019), it is unclear to what extent transparency of AI's decision-making rules mitigates uncertainty and enhances trust and whether such transparency can compensate for the loss of human agency in machine-learning AI's decision-making. It is also unclear how users process AI's explanations. Do they care to scrutinize the explanations? Does the quality of the explanations matter? Answers to these questions will shed light on users' information processing during human–AI interaction and provide suggestions for XAI designers and policymakers. To address these questions, this study investigates the effects of the type of AI's decision-making rules (i.e., human-made vs. machine-generated) and transparency of these rules on users' uncertainty about, trust in, and intention to use decision-making AI across two common AI applications—fake news detection and personality assessment.

Literature review

Uncertainty reduction and trust in human–machine interaction

Concepts of trust and uncertainty reduction were initially studied in the context of human–human (H–H) interaction but later were also found applicable to human–machine (H–M) interaction (Jian, Bisantz, & Drury, 2000). In both contexts, trust denotes "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 51). For users to comply with AI-made decisions or to accept an AI system making decisions on behalf of them, they first need to trust it. Given that machines are typically considered as lacking intentionality (Dennett, 1988), this study focuses on the *competence* dimension of individuals' trust in decision-making AI systems (Lee & Moray, 1992).

How does trust develop? Research suggests that uncertainty reduction facilitates goal achievement, and therefore, is fundamental to trust development in both H–H and H–M contexts (Jian et al., 2000). Uncertainty—defined as one's lack of ability to predict and explain the behaviors of their interaction partner (Berger, 1987; Clatterbuck, 1979)—arises when individuals lack sufficient predicative or explanatory knowledge about the other to accurately predict interaction outcomes (Berger & Bradac, 1982). Because perfect knowledge about each other is impossible, all interaction episodes involve uncertainty, even in long-term relationships (Berger, 1987). However, according to uncertainty reduction theory (URT), uncertainty is especially substantial when two strangers just meet (Berger & Calabrese, 1975). During the initial stage of their interaction, interactive pairs know little about each other's attitudes, beliefs, and qualities, and therefore cannot accurately predict or explain each other's behaviors (Berger & Calabrese, 1975). Given the novelty of decision-making AI, most interactions between individuals and AI systems may involve such new meetings, in which individuals have little experience with their machine partners and thereby may endure high uncertainty as a result.

URT also suggests that individuals are inherently motivated to reduce uncertainty by gathering information that could help them predict and explain each other; such information allows them to make better plans to achieve their goals during and through communication (Berger & Calabrese, 1975). In other words, uncertainty reduction is driven by an "effectance motivation," that is, the motivation to be a competent agent in one's environment, which is a basic human motivation (White, 1959, p. 321). As such, although URT was originally developed in the context of H–H interaction, uncertainty reduction as a goal-driven process should also apply to other contexts in which the interactants can potentially influence one's outcomes (Berger, 1987). Because AI-made decisions, once adopted, will influence users' outcomes, users are expected to be motivated to reduce uncertainty about the decision-making AI system.

Uncertainty reduction is essential for developing trust with machines. The research found that predictive knowledge and explanatory knowledge about an automated system (e.g., clear explanations of how the machine works) enhanced users' trust in it (e.g., Jian et al., 2000). Even when the system erred, users still trusted it as long as they were given the logic explaining why the error might have occurred (Dzindolet, Peterson, Pomranky, Pierce, & Beck, 2003). For a decision-making AI system, its decisions can be explained and predicted by the rules it follows, knowing which may reduce users' uncertainty and enhance trust in it. The use of machine-learning techniques, however, makes it difficult for users to access AI's decision-making rules. Unlike traditional AI systems that are programmed to follow human-made rules, machine-learning AI systems learn rules from data. These learned rules are often unintelligible and might be fundamentally different from human logic. This difference between these two types of AI, defined as a shift in AI's agency locus from human to machine, may influence users' perceptions of AI's decision-making.

Defining agency locus of machines

To define the agency locus of machines, we first reviewed two related concepts, agency and apparent agency. *Agency*, if broadly defined, denotes the exercise of the capacity to act; if narrowly defined, it only counts intentional actions (Schlosser, 2019). According to the narrow definition, machines do not have agency because they lack intentionality (Dennett, 1988). So, the term "machine agency," when used, often refers to machines' *apparent agency*. Apparent agency denotes machines' exercise of thinking and acting capacities that they appear to have during H–M interaction, as opposed to the truth upon users' reflection—machines are programmed by humans to act in specific ways and have no real agency (Takayama, 2015). That said, not all machines are fully programmed by humans to follow human-made rules.

Modern AI systems built on machine-learning techniques take a different approach. Machine learning, defined as "the capacity to define or modify decision-making rules autonomously" (Mittelstadt et al., 2016, p. 3), grants an AI system some degree of autonomy and makes it less subject to human determination as compared with fully programmed machines. This shift toward automated learning disputes the early view of machines as merely a proxy for its human creators (e.g., Dennett, 1988) for two reasons: First, machine-learning AI is not fully programmed with human-made rules; and second, even its programmers do not understand what the machine-learning algorithms have learned from the data (Mittelstadt et al., 2016).

To describe this difference between fully programmed AI and machine-learning AI, we introduce the concept of *agency locus*. Just as with the concept "locus of control"—in which "locus" refers to the cause that a person attributes their success or failure to (i.e., internal or external; Rotter, 1966)—we define the locus of a machine's agency as its rules, which are the cause of its apparent agency and can be "external" (e.g., created by humans) or "internal" (e.g., generated by the machine). Accordingly, the agency locus of preprogrammed AI systems is human-made rules and therefore reflects human agency, whereas the agency locus of machine-learning AI systems is machine-generated rules, which reflect certain machine agency. The two types of AI systems are, respectively, labeled as "human-agency AI" and "machine-agency AI."

Effects of agency locus on uncertainty and trust

Just as the initial stage of H–H interaction is characterized by high uncertainty (Berger & Calabrese, 1975), human users may also experience uncertainty at the initial stage of their interaction with an AI system. According to the three-factor theory of anthropomorphism (Epley et al., 2007), the need to reduce uncertainty about a nonhuman agent motivates individuals to see the humanness in it, that is, to experience *social presence*. Social presence refers to a psychological state in which individuals experience a para-authentic object (e.g., a game avatar) or an artificial object (e.g., a computer) as humanlike, with or without their conscious awareness (Lee, 2004). Researchers of H–M interaction have found social presence to be a stable phenomenon (Oh, Bailenson, & Welch, 2018; Reeves & Nass, 1996), which can be triggered by a variety of anthropomorphic cues embedded in machines (e.g., avatars, computers, and robots), such as humanlike physical attributes (e.g., face), human-like behaviors (e.g., gesture and interactivity), and social signals (e.g., social identities and roles; Lombard & Xu, 2021).

Research suggests that perceiving machines as humanlike (i.e., experiencing social presence) may help users reduce uncertainty by facilitating simulation of human intelligence, which is a familiar type of intelligence to human users. For example, the research found a human-like robot triggered a richer and more anthropomorphic mental model among users than a machine-like robot (Kiesler & Goetz, 2002). Users also claimed more knowledge about a human-like robot than a machine-like one (Ososky, Philips, Schuster, & Jentsch, 2013). In addition, students' self-reported social presence with a robot teacher was found negatively associated with uncertainty (Edwards, Edwards, Spence, & Westerman, 2016). As such, should users perceive the decision-making AI system as human-like, they may apply their anthropocentric knowledge about how a typical human makes decisions to understand and predict the AI system's decision-making, thereby reducing uncertainty and increasing trust.

Agency locus being human-made rules might function as an anthropomorphic cue and trigger social presence. A recent study found that algorithmic human likeness (i.e., algorithms profiling users in ways similar to humans' impression formation) induced more social presence and positive attitudes among users (Liu & Wei, 2021). Since human-agency AI follows the rules that reflect collective human intelligence crystallized from humans' experiences and reflections, its human likeness at the algorithmic level may facilitate users to perceive it as humanlike. In contrast, machine-agency AI follows self-learned rules that might be fundamentally different from human logic. Therefore, machine-agency AI may induce less social presence and lead to higher uncertainty and lower trust than human-agency AI, *ceteris paribus*. Specifically, trust was assessed in terms of both attitudes and behavioral intention with three measures: (a) trust in the AI-made judgments (i.e., "trust in judgments"); (b) trust in the AI system's competence (i.e., "trust in system"); and (c) intention to use the system in the future (i.e., "use intention").

H1: Compared with human-agency AI, machine-agency AI induces less social presence.
H2: Social presence mediates the positive effect of machine-agency AI (vs. human-agency AI) on uncertainty.
H3: Social presence and uncertainty mediate the negative effects of machine-agency AI (vs. human-agency AI) on (a) trust in judgments, (b) trust in system, and (c) use intention.

Transparency as a remedy: effects of transparency on uncertainty reduction and trust

If users distrust AI because of their uncertainty about it, will transparency—telling users the rationales for its decisions—alleviate it? As explanatory knowledge is the most direct way to reduce uncertainty about another (Berger & Bradac, 1982), transparency of a decision-making AI system is expected to reduce uncertainty, and as a result, enhance users' trust in it.

H4: Compared with no transparency, transparency of AI decreases uncertainty.
H5: Uncertainty (reduction) mediates the positive effect of transparency (vs. no transparency) on (a) trust in judgments, (b) trust in system, and (c) use intention.

Moreover, the availability of such individual-based information about an AI system may also interfere with the effects of agency locus as an anthropomorphic cue. When users draw on anthropocentric cognitions to reduce uncertainty, they rely on their knowledge about humans as a category (i.e., knowledge about how a typical human makes decisions), which is not specific enough to enable accurate predications or explanations. For example, knowing that an AI system follows human-made rules does not explain why it judges someone as hostile. Without rule transparency, users can only rely on category-based information to understand an AI system. But when the individual-based information about the AI system is available, category-based knowledge might play a lesser role in shaping users' perceptions, as suggested by research on impression formation (Fiske & Neuberg, 1990). Thus, transparency is expected not only to reduce uncertainty and enhance trust overall but also to attenuate the effects of agency locus as an anthropomorphic cue (i.e., its indirect effects via social presence on uncertainty and trust).

H6: Transparency moderates the indirect effect of agency locus on uncertainty via social presence, with the indirect effect being more prominent in the no-transparency condition.
H7: Transparency moderates the indirect effects of agency locus on (a) trust in judgments, (b) trust in system, and (c) use intention via social presence and uncertainty, with these indirect effects being more prominent in the no-transparency condition.

Is transparency a heuristic cue? Exploring how users process explanatory messages

Hypotheses 4–7 are based on the informational value of AI transparency, such that transparency provides valuable information about the AI system, allowing users to predict and explain AI-made decisions more accurately. An alternative explanation, however, is that transparency may merely function as a heuristic cue (e.g., Shin & Park, 2019). Because transparency is a positive feature for AI to have, users may rate an AI system more positively if they see it provides explanations for its decisions, even without carefully examining the explanations' quality.

Theories and research in information processing suggest that it is likely that transparency acts as a heuristic cue. Dual-process theories (e.g., elaboration likelihood model and heuristic-systematic model) suggest that individuals process information in two different ways: One process is characterized by deep, effortful, and systematic analyses of stimuli, while the other entails automatic, shallow, and mindless processing (Eagly & Chaiken, 1984). As "cognitive misers" (Fiske & Taylor, 1984, p. 141), individuals only engage in the effortful processing when they are motivated (e.g., when the messages are on a topic that they care about). When not motivated, they rely on easy-to-process cues in a message (e.g., length, number of arguments) and heuristics (i.e., rules of thumb, such as "If an argument is long, then it must be solid.") to form impressions. In this case, the mere presence of explanations may serve as a heuristic cue that triggers positive perceptions; the quality (i.e., the informational value) of the explanations may matter less, if at all. For example, Langer, Blank, and Chanowitz (1978) found that people complied with requests (i.e., cutting the line and making copies) accompanied by a placebic explanation that was tautological to the requests (i.e., "because I have to make copies," p. 637) as much as requests followed by a real, solid explanation (i.e., "because I'm in a rush," p. 637), especially when the request was smaller (i.e., printing 5 pages vs. 20 pages). They concluded that when individuals are "mindless," not the content of an explanation but its explanatory *form* is persuasive.

Are AI users also mindless when interacting with AI systems? Do they care enough to process the content of AI's explanations thoroughly? To explore whether the effects of transparency on uncertainty and trust are merely due to its symbolic value as a heuristic cue, this study included a condition of placebic transparency, following the design of the copy machine study by Langer et al. (1978). Specifically, placebic transparency only has the *form* of an explanation (i.e., "because. . .") and is tautological to the AI-made decision that it is supposed to explain (e.g., "The AI system judges this news article as fake, because the news article is identified as not real."). Thus, it provides no information about the rationales for AI's decision-making. We accordingly relabel the "transparency" condition in hypotheses 4–7 as "real transparency," for it has real informational value. If placebic transparency reduces uncertainty and enhances trust, it suggests that users might have treated AI transparency as a heuristic cue.

In addition, the agency locus of AI might also affect users' motivations to process AI's explanations. As users may feel less uncertainty about the human-agency AI (vs. machine-agency AI), users of human-agency AI may have a lower need to reduce uncertainty further, and therefore a lower motivation to scrutinize AI's explanations. In this case, placebic transparency may pass for its explanatory form. Hence, we propose first to explore the overall effects of placebic transparency and then to examine how agency locus may influence its effects, if any.

RQ1: Compared with no transparency, what are the effects of placebic transparency on (a) uncertainty, (b) trust in judgments, (c) trust in system, and (d) use intention?

H8: The effects of placebic transparency (vs. no transparency) on (a) uncertainty, (b) trust in judgments, (c) trust in system credibility, and (d) use intention are more prominent for human-agency AI than for machine-agency AI.

Method

To enhance this study's external validity, we tested the hypotheses in contexts of two tasks that may require different skills: detecting fake news (i.e., AI making factual judgments, which involves "mechanical" skills) and assessing social media users' agreeableness (i.e., AI making social judgments, which involves "human" skills; Lee, 2018). In a 2 (agency locus: human-made rules vs. machine-learned rules) × 3 (transparency: no vs. placebic vs. real) × 2 (task: fake news detection vs. agreeableness assessment) between-subjects online experiment, participants first filled out a questionnaire measuring their knowledge and attitudes related to AI, then were randomly assigned to interact with a website that was purported to host a decision-making AI system, and then answered questions about their experience with the AI system.

Participants

Participants ($N = 491$) were recruited from a Qualtrics panel through quota sampling, with distributions of gender ($n_{female} = 252$, 51%) and age ($M = 46.8$ years, $Mdn = 48$, $SD = 17.0$, range: 18–83) matched to those of the U.S. Census Bureau's 2017 American Community Survey (United States Census Bureau, 2018). The majority of the participants were Caucasians (83%), followed by African Americans (8%), Hispanics/Latinos (6%), Asians (5%), Native Americans (2%), and multiracial (under 1%). About half of the participants had a bachelor's degree or above (44%) and an annual household income of over $50,000 (49%).

Stimuli

Participants were told that they could download the AI system as a browser extension after experiencing its functions on a website. The fake news detection AI *DeepTruth* was introduced as an AI system that could detect and remove fake news from users' news feeds. The personality assessment AI *SmartFriending* purportedly could identify disagreeable social media users based on their online activities and remove them from one's network. On the website, participants saw the AI system making four binary judgments one at a time on four news stories or four *Twitter* user profiles purported to be retrieved online in real time. But in reality, all materials were set up on the websites beforehand without using any actual AI technology.

Manipulation of agency locus

We manipulated the agency locus on the website's landing page by telling participants how the rules governing the system's judgments were developed, using both animated text and visual illustration (see Figure 1). For *DeepTruth*, participants in the human-agency [machine-agency] condition read (a) "Human experts [*DeepTruth*'s algorithms] analyzed data of fake and real news," (b) "Human experts [*DeepTruth*] identified [learned] patterns of fake news based on their [its] analyses," and (c) "*DeepTruth* applies these rules human experts have identified [it has learned] in judging the authenticity of online news." For *SmartFriending*, similarly, participants in the human-agency [machine-agency] condition read (a) "Human experts [*SmartFriending*'s algorithms] analyzed data of social media users' online behaviors and personalities," (b) "Human experts [*SmartFriending*] identified [learned] patterns of (dis)agreeable users from their [its] analyses," and (c) "*SmartFriending* applies these rules human experts have identified [it has learned] in judging the agreeableness of a social media user."

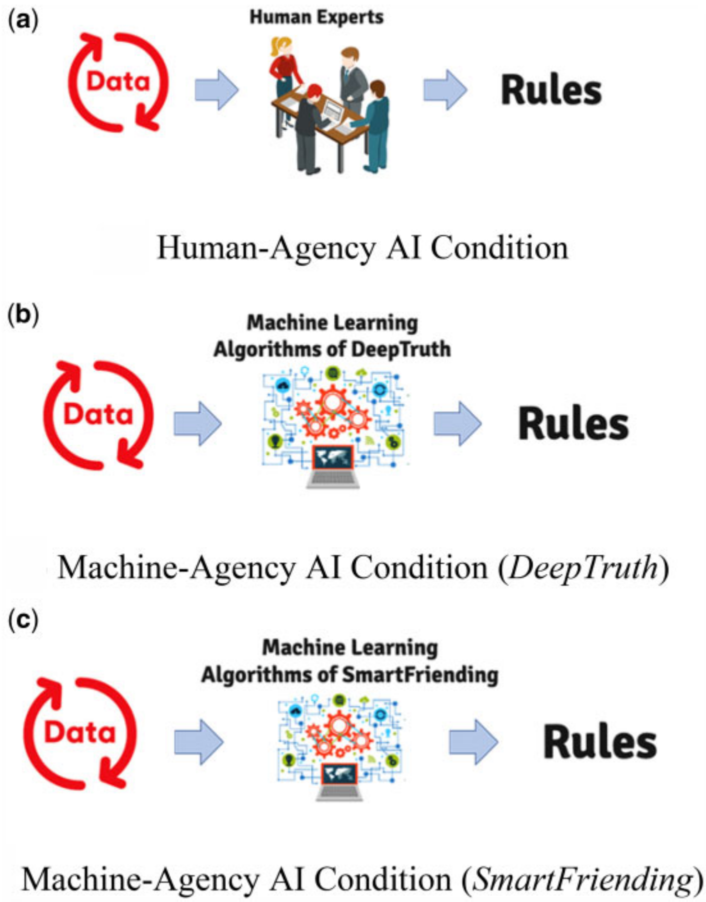

Figure 1 Visual cues indicating agency locus. (a) Human-agency AI condition, (b) machine-agency AI condition (*DeepTruth*), and (c) machine-agency AI condition (*SmartFriending*).

Manipulation of transparency

Following each judgment (e.g., “*DeepTruth* judges it as True.” or “*SmartFriending* judges the user as Disagreeable.”), transparency was manipulated (see Figure 2). In the no-transparency condition, no explanations were provided. In the placebic-transparency condition, three explanations tautological to the judgment were provided, such as “*DeepTruth* judges it as True… because the news story is determined as not fake,” or “*SmartFriending* judges the user as Disagreeable… because this user is identified as not amiable.” In the real-transparency condition, three solid explanations were provided, such as “*DeepTruth* judges it as Fake… because contents in this news story are at odds with other news sources,” or “*SmartFriending* judges the user as Agreeable… because this user has frequently expressed positive emotions in tweets.” All real explanations were generated based on research in fake news and personality perceptions and were validated as high-quality explanations through a pretest (see Supporting Information).

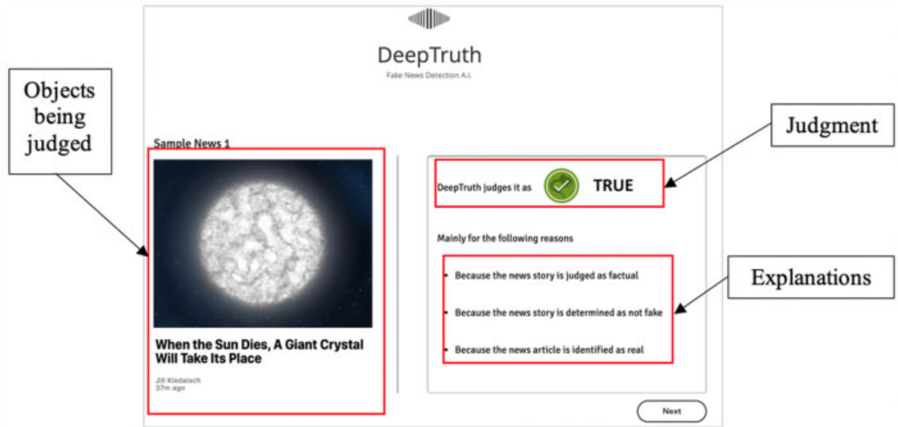

Figure 2 Manipulation of transparency.

Targets and judgments

Within each condition, four variations of the AI system (labeled as -A1, -A2, -B1, -B2) were created to rule out the effects of the targets to be judged (i.e., news stories and *Twitter* users) and the valence of the judgments. *DeepTruth*-A1 judged the first two news stories as "True" and the last two as "Fake," *DeepTruth*-A2 evaluated the same news stories but judged the first two as "Fake" and the last two as "True." *DeepTruth*-B1 and *DeepTruth*-B2 evaluated four different news stories in which judgments were also counterbalanced. These eight news stories (on topics of technology, astronomy, military, and climate) were rated as the most ambiguous in terms of their "fakeness" in a pretest. Four variations of *SmartFriending* were created in the same way. The gender and level of customization of the eight *Twitter* user profiles were balanced across the two sets. These variations resulted in a total of 48 different stimuli websites.

Measurement

All variables were measured on seven-point scales. Details are presented in Table 1.

Results

Manipulation check

Manipulation of agency locus was successful. Results of two independent samples t-tests show that the rules of human-agency AI were perceived as more human-made ($M = 4.7$, $SD = 2.1$) than machine-agency AI ($M = 2.9$, $SD = 1.8$), Cohen's $d = 0.89$, $t(480) = 9.87$, $p < .001$; human-agency AI was also perceived as more dependent on human programming ($M = 5.5$, $SD = 1.5$) than machine-agency AI ($M = 4.9$, $SD = 1.6$), Cohen's $d = 0.42$, $t(489) = 4.68$, $p < .001$.

Manipulation of transparency was also successful based on the results of two one-way analysis of variance. On the item "*DeepTruth/SmartFriending* revealed the reasons for its judgments," the real-transparency condition received the highest rating of agreement ($M = 6.2$, $SD = 1.2$), significantly higher than the placebic-transparency condition ($M = 5.5$, $SD = 1.8$); and both were significantly higher than the no-transparency condition ($M = 2.9$, $SD = 2.0$), $ps < .017$ (with Bonferroni

Table 1 Measures of variables

Variables	No. of items	Sample items	Cronbach's α
Agency locus (manipulation check)		1. The rules for detecting fake news/judging one's agreeableness were devised by *machine/human* 2. How much do you think the system's functioning relies on human programming?	
Transparency (manipulation check)		1. The system revealed the reasons for its judgments 2. The system's rationales for its judgments were sound	
Social presence (Lee, Peng, Jin, & Yan, 2006)	6	When receiving the judgments, how much did you feel • as if you were interacting with a person? • as if a person was thinking behind the screen?	0.97
Uncertainty (Clatterbuck, 1979)	6	How confident were you in predicting the system's judgments? How certain were you about what this system is really like?	0.93
Trust in judgments (Lee & See, 2004)	9	To what extent do you think the judgments are____ Correct; Accurate	0.98
Trust in system (McCroskey & Teven, 1999)	12	Please rate *DeepTruth/SmartFriending* on: • *unintelligent/intelligent* • *incompetent/competent*	0.98
Use intention (Kim & Sundar, 2014)	4	I would recommend others to use *DeepTruth/SmartFriending* • I think I will use *DeepTruth/SmartFriending* in the future	0.97
Anthropomorphism (Sundar, 2004)	7	If the computer does not perform the function I ask, I will curse it I would be glad if a computer praised my job	0.87
Need for cognition (Lins de Holanda	6	I would prefer complex to simple problems	0.88

(*Continued*)

Table 1 (continued)

Variables	No. of items	Sample items	Cronbach's α
Coelho, Hanel, & Wolf, 2020)		Thinking is not my idea of fun. (Reverse coded)	
Awareness of AI use	8	How likely is that the following technologies are using AI? Virtual assistants; Smart speakers	0.90
AI acceptance	17	It is acceptable to have AI • make decisions with real-world consequences for humans • decide consumers' personal finance score	0.94
General trust in AI	4	If a decision is made by AI, it must be precise If a decision is made by AI, it must be a good decision	0.93
Perceived AI's threat	7	I am worried that AI would threaten human dignity I am concerned that AI would threaten human society	0.93
Issue familiarity (four variables)		How familiar are you with the following issues? a. Technology; b. Astronomy; c. Military; d. Climate	
Social media experience	6	Using social media is part of my daily routine I am familiar with the features of mainstream social media	.92

Note: All variables were measured on 7-point scales.

correction), $F(2, 488) = 182.99$, $p < .001$. On perceived soundness of the rationales, the difference between the placebic-transparency ($M = 4.4$, $SD = 2.0$) and no-transparency condition ($M = 3.9$, $SD = 1.8$) was not significant, $p = .097$; both were rated lower than the real-transparency condition ($M = 5.3$, $SD = 1.8$), p's $< .001$, $F(2, 488) = 21.82$, $p < .001$.

Hypotheses testing

Hypotheses were tested with PROCESS Macro (v3.4, Hayes, 2018), with 5,000 bootstraps resamples and 95% percentile bootstrap confidence intervals (CIs). Seven dummy variables representing the eight stimulus versions were controlled for as covariates (versions within each task were equivalent). Another 11 variables of individual differences that might influence the outcome variables were also controlled to boost data analyses' power, which might have been attenuated by the heterogeneity of the sample (Shadish, Cook, & Campbell, 2002). Because one's experience in a task domain may influence their perceptions of AI's judgments in that area, participants' experiences related to the stimuli were controlled, including their familiarity with the four news topics (i.e., technology, astronomy, military, and climate) and experience with social media. The other six variables included one's anthropomorphism tendency, awareness of AI use, acceptance of AI, general trust in AI, perceived AI's threat, and need for cognition, which is expected to influence users' experience with AI systems based on past research (e.g., Araujo et al., 2020).

Hypotheses 1 and 2 were tested with Model 4, with agency locus (1 = machine-agency AI) as the independent variable, social presence as the mediator, and uncertainty as the dependent variable; both were supported. Machine-agency AI induced lower social presence (hypothesis 1), $B = -0.33$, $SE = 0.13$, $p = .009$, 95% CI [−0.58, −0.08], Cohen's $d = -0.20$. Its indirect effect on uncertainty via social presence was significant (hypothesis 2), $B = 0.08$, $BootSE = 0.03$, 95% CI [0.02, 0.15], $ab_{cs} = .05$.

Hypothesis 3, tested with Model 6, was also supported. Social presence and uncertainty mediated the effect of agency locus on trust in judgments (hypothesis 3a), $B = -0.04$, $BootSE = 0.02$, 95% CI [−0.07, −0.01], $ab_{cs} = -.02$, trust in system (hypothesis 3b), $B = -0.03$, $BootSE = 0.01$, 95% CI [−0.06, −0.01], $ab_{cs} = -.02$, and use intention (hypothesis 3c), $B = -0.03$, $BootSE = 0.02$, 95% CI [−0.07, −0.01], $ab_{cs} = -.02$.

Hypotheses 4 and 5, tested with Model 4, were supported. Real transparency reduced uncertainty (hypothesis 4), $B = -0.90$, $SE = 0.13$, $p < .001$, 95% CI [−1.16, −0.64], Cohen's $d = -0.64$. Its indirect effect via uncertainty was significant on trust in judgments (hypothesis 5a), $B = 0.42$, $BootSE = 0.08$, 95% CI [0.28, 0.58], $ab_{cs} = .26$, trust in system (hypothesis 5b), $B = 0.36$, $BootSE = 0.07$, 95% CI [0.23, 0.52], $ab_{cs} = .24$, and use intention (hypothesis 5c), $B = 0.45$, $BootSE = 0.09$, 95% CI [0.29, 0.64], $ab_{cs} = .24$.

Hypotheses 6 and 7 that real transparency moderates the indirect effects of agency locus were tested with Models 14 and 91, respectively. In support of both, the indirect effects of agency locus on uncertainty and trust were only significant in the no-transparency condition (see Table 2).

RQ1 was tested with Model 4. The effect of placebic transparency was not significant on uncertainty (RQ1a), $B = -0.24$, $SE = 0.14$, $p = .092$, 95% CI [−0.51, 0.04], Cohen's $d = -0.16$, trust in judgments (RQ1b), $B = 0.23$, $SE = 0.14$, $p = .117$, 95% CI [−0.06, 0.51], Cohen's $d = 0.15$, and trust in system (RQ1c), $B = 0.20$, $SE = 0.15$, $p = .164$, 95% CI [−0.08, 0.49], Cohen's $d = 0.13$. But its effect on use intention was positive (RQ1d), $B = 0.42$, $SE = 0.17$, $p = .015$, 95% CI [0.08, 0.75], Cohen's $d = 0.22$. Hypothesis 8 was tested with Model 1. Hypotheses 8a–c were supported such that placebic transparency reduced uncertainty, increased trust in judgments, and increased trust in system but only for human-agency AI; hypothesis 8d was not supported (see Table 2 for details).

Discussion

Machine-learning techniques grant AI more autonomy while posing greater challenges for human users to understand and trust its decision-making. Through comparing individuals' perceptions of two types of AI systems (human-programmed vs. machine-learning) and examining the effects of transparency, findings illuminate the processes of uncertainty reduction and trust development in humans' interactions with decision-making AI systems.

The first hypothesis stated that machine-agency AI would induce less social presence than human-agency AI. Results support this hypothesis and suggest that lay users do care about such an

Table 2 Moderated mediation and moderation tested with PROCESS (v3.4)→

IV → DV	Med/Mod	Statistics
H6: Agency locus → Uncertainty	Model 14 Med: Social presence Mod: Real transparency	Index = −0.10, *BootSE* = 0.04, 95% CI [−0.19, −0.02] No: $B = 0.11$, *BootSE* = 0.04, 95% CI [0.03, 0.20], $ab_{cs} = 0.07$ Real: $B = 0.01$, *BootSE* = 0.02, 95% CI [−0.04, 0.06], $ab_{cs} = 0.004$
H7a: Agency locus → Trust in judgments	Model 91 Med: Social presence, uncertainty Mod: Real transparency	Index = 0.05, *BootSE* = 0.02, 95% CI [0.01, 0.10] No: $B = -0.05$, *BootSE* = 0.02, 95% CI [−0.10, −0.01], $ab_{cs} = -0.03$ Real: $B = -0.003$, *BootSE* = 0.01, 95% CI [−0.03, 0.02], $ab_{cs} = -0.002$
H7b: Agency locus → Trust in system		Index = 0.04, *BootSE* = 0.02, 95% CI [0.01, 0.08] No: $B = -0.04$, *BootSE* = 0.02, 95% CI [−0.08, −0.01], $ab_{cs} = -0.03$ Real: $B = -0.003$, *BootSE* = 0.01, 95% CI [−0.02, 0.02], $ab_{cs} = -0.002$
H7c: Agency locus → Use intention		Index = 0.05, *BootSE* = 0.02, 95% CI [0.01, 0.10] No: $B = -0.05$, *BootSE* = 0.02, 95% CI [−0.10, −0.01], $ab_{cs} = -0.03$ Real: $B = -0.003$, *BootSE* = 0.01, 95% CI [−0.03, 0.02], $ab_{cs} = -0.002$
H8a: Placebic → Uncertainty	Model 1 Mod: Agency locus	Interaction term: $\Delta R^2 = .005$, $F(1, 468) = 3.57$, $p = .059$ Human-agency: $B = -0.46$, $SE = 0.18$, $p = .012$, 95% CI [−0.81, −0.10], Cohen's $d = -0.31$ Machine-agency: $B = -0.01$, $SE = 0.18$, $p = .949$, 95% CI [−0.37, 0.35], Cohen's $d = -0.01$
H8b: Placebic → Trust in judgments		Interaction term: $\Delta R^2 = 0.01$, $F(1, 468) = 9.30$, $p = .002$ Human-agency: $B = 0.59$, $SE = 0.18$, $p = .002$, 95% CI [0.23, 0.95], Cohen's $d = 0.38$ Machine-agency: $B = -0.14$, $SE = 0.19$, $p = .446$, 95% CI [−0.51, 0.23], Cohen's $d = -0.09$
H8c: Placebic → Trust in system		Interaction term: $\Delta R^2 = 0.004$, $F(1, 468) = 3.13$, $p = .078$ Human-agency: $B = 0.42$, $SE = 0.19$, $p = .025$, 95% CI [0.05, 0.79], Cohen's $d = 0.28$ Machine-agency: $B = -0.01$, $SE = 0.19$, $p = .954$, 95% CI [−0.39, 0.36], Cohen's $d = -0.01$

(*Continued*)

Table 2 (continued)

IV → DV	Med/Mod	Statistics
H8d: Placebic → Use intention		Interaction term: $\Delta R^2 = .0004$, $F(1, 468) = 0.35$, $p = .557$

Note: Machine-agency AI coded as 1.
ab_{cs} = completely standardized indirect effect; CI = confidence interval; DV = dependent variable; Index = index of moderated mediation; IV = independent variable; Med = mediator; Mod = moderator; No = no transparency; Placebic = placebic transparency; Real = real transparency.

algorithmic-level difference when made aware of it. This finding is consistent with previous research findings on agency's effects in the virtual environment, where participants interacted with a virtual character that was purported to be controlled either by a human or by a computer when both characters looked and behaved the same. Meta-analyses show that individuals experienced more social presence with and showed more social responses (e.g., maintaining a larger physical distance) to the human-controlled avatar than the computer-controlled agent (Fox et al., 2015; Oh et al., 2018). Although their agency manipulation differed from agency locus in the present study (which did not involve direct human control of any system), both findings suggest that individuals do care about the origin of the machine's apparent agency and psychologically orient to that origin during the interaction when made aware of it.

Moreover, social presence was associated with lower uncertainty and higher trust as hypotheses 2 and 3 predicted. This finding suggests that people not only orient to the agency locus of an AI system but also use it as an informational cue to reduce uncertainty, if applicable. When participants interacted with the human-agency AI system, knowing that it was programmed to follow human-made rules seemed to have allowed them to simulate the "mental states" of the AI system and led them to feel more knowledgeable about its decision-making process. This finding is also consistent with previous research that found anthropomorphic attributes of machines made people feel more knowledgeable and less uncertain about the machines (e.g., Kiesler & Goetz, 2002; Ososky et al., 2013). To be noted, although anthropomorphism was found conducive to trust development in this study, this might only be the case when machines are at the low end of human likeness; a highly but imperfectly human-like interface may cause negative evaluations and emotions, as suggested by the uncanny valley hypothesis (Mori, MacDorman, & Kageki, 2012).

We also explored to what extent the disclosure of rationales for AI's decisions could mitigate the uncertainty and distrust people have with decision-making AI. Results confirmed our expectations (hypotheses 4 and 5) that real transparency can reduce uncertainty and enhance users' trust in decision-making AI, consistent with findings in previous research (e.g., Shin & Park, 2019). Moreover, real transparency attenuated the effects of agency locus as an anthropomorphic cue on uncertainty reduction and trust as hypotheses 6 and 7 predicted, suggesting that transparency can remedy the consequences of the loss of humanness in machine-learning AI.

To find out whether the effects of AI transparency were merely the results of heuristic processing, we examined the effects of placebic transparency on uncertainty and trust (RQ1) and expected that if there was any effect, it would be more prominent for the human-agency AI (hypothesis 8). We found that placebic transparency reduced uncertainty and enhanced trust for the human-agency AI (but not for the machine-agency AI), and increased use intention across both types of AI systems. First, this finding supports the hypothesis that users feel more uncertain about the machine-agency AI, in which case only high-quality explanations could address users' higher need for uncertainty reduction. Second, in light of dual-process theories, which posit that people only scrutinize the messages when they are motivated (Eagly & Chaiken, 1984), this finding suggests that users are not always motivated to process AI's explanations carefully. Otherwise, participants would have noticed the lack of informational value in placebic transparency and would not have been influenced. Thus, we speculate that the observed effects of AI transparency in this study and previous research might be partially attributed to transparency's symbolic value.

Theoretical implications

These findings also have theoretical implications for understanding H–M interaction. First, our findings advance the explanatory power of uncertainty reduction theories originally developed for H–H interaction. For example, the Anxiety/Uncertainty Management (AUM) theory, developed based on URT, posits that communication with outgroup members is characterized by high uncertainty; without person-based information, individuals draw on group-based information to manage uncertainty; intergroup similarity reduces uncertainty; and with person-based information available, group-based information will be less relied upon because person-based information increases prediction accuracy (Gudykunst, 1995).

The process of uncertainty reduction observed in this study seems to resemble that in H–H interaction. AI, as a new form of intelligence, is essentially an outgroup member whom human users know little about. At the initial stage of interaction, as in the current study, users drew on group-based information (i.e., agency locus) to reduce uncertainty about the AI system; with higher intergroup similarity (i.e., programmed by humans), users felt less uncertainty; and when individual-based information (i.e., explanations for AI's decisions) was available, the effects of group-based information were attenuated. As such, uncertainty reduction theories such as URT and AUM may be extended to explain and predict the processes and outcomes of H–M interaction.

Second, findings suggest a boundary condition for the phenomenon of "Computers Are Social Actors" (CASA)—individuals responding socially to machines by applying social scripts developed from H–H interaction (Reeves & Nass, 1996). Our findings suggest that the application of H–H social scripts in H–M interaction might be conditional on the unavailability of information that allows individuals to accurately explain and predict the machines' behaviors (e.g., information about their action rules). In many CASA studies, participants were led to interact with zero-history machines and were never given individuated information about those machines that could help reduce uncertainty (Reeves & Nass, 1996). As such, participants might endure high uncertainty and be therefore motivated to simulate human intelligence to reduce uncertainty (Epley et al., 2007) and treated machines like persons as a result. Should they have individuated information about the machines, as in the real-transparency condition in this study, they could generate more accurate predictions and explanations of the machines' behaviors and need not simulate human intelligence or draw on the learned social scripts to reduce uncertainty.

Third, our findings suggest a more nuanced picture of source orientation in H–M interaction than the dichotomy between "Computer-as-Source" (CAS) versus "Computer-as-Medium" models (CAM; Sundar & Nass, 2000). The CAS model suggests that users treat a computer as an independent source during interaction without thinking about the source of its agency (i.e., its programmer). Alternatively, the CAM model suggests that users orient to the computer programmer and treat the computer as merely a medium. Sundar and Nass (2000, Study 1) found support for the CAS model, as participants evaluated the two computers labeled as either "computer" or "programmer" differently, which suggests that they did not orient to the computer programmer; otherwise, they would have rated the two computers similarly. Findings in this study, however, suggest that users treated the AI system at least partially as a medium and oriented to its distal source of agency (i.e., its rules). Otherwise, no effects of agency locus would have occurred. Although this seems to contradict the conclusion from Sundar and Nass (2000), they are not incompatible if the findings in Sundar and Nass (2000) are reinterpreted without dichotomizing the two models. Between CAS and CAM, there could be a middle ground, in which participants in the "computer" condition might still orient to the programmer (i.e., as CAM suggests) but to a lesser degree than those in the "programmer" condition, who thought they were interacting with the programmer more directly. Such a middle ground can be useful to model humans' interaction with emerging AI-based machines, whose agency is increasingly complicated by the participation of users, designers, and institutions in building and operating them. In such cases, conceptualizing machine agency as a chain, or even a network, may better characterize users' agency attribution and source orientation during H–M interaction.

Practical implications

Knowing how individuals perceive decision-making AI also informs how AI should be designed and regulated. Based on the current findings, first, AI interface designers may consider disclosing the involvement of human agency in building an AI system. Findings suggest this disclosure can help establish a common ground with users and facilitate uncertainty reduction and trust development. Second, given the effects of transparency as a heuristic cue, policymakers may consider regulating the quality of XAI to protect the public from being exploited. Accordingly, XAI designers should consider both the informational value of AI's explanations and users' motivations to process them. In response, they should design easy-to-process and informative explanations to enhance user experience and increase use intention.

Limitations and future directions

In evaluating the results of this study, some limitations need to be noted. First, the tasks of AI systems—screening fake news and (un)following social media users—are relatively low-stakes, meaning user involvement might be low in the current study. Future research may explore the effects of agency locus and transparency of AI in high-stakes scenarios such as when AI making decisions about hiring, resource allocation, and even (in the case of self-driving cars) whom (not) to kill and may formally examine the effects of users' involvement.

Second, agency locus's effect sizes were small overall, except for its effect on social presence. This limited influence could be due to low user involvement and the subtleness of the operationalization of agency locus. Another reason may be the multi-determinedness of H–M interaction. Various aspects of technology (e.g., appearance and interactivity) may interact with or even override the effects of agency locus. Future research may use stronger manipulation and further explore the combined effects of an AI system's appearance, performance, and agency locus on user experience and trust.

Third, although agency locus was found to influence social presence, ratings of social presence were low for both the human-agency AI ($M = 3.1$, $SD = 1.7$) and the machine-agency AI ($M = 2.8$, $SD = 1.6$). This could be due to the low human likeness of the AI systems' interface and the task-oriented nature of the current contexts. Participants might have focused only on the cognitive capacities of the AI system and did not simulate a full-blown person. Future research may consider refining the measure of social presence to capture the perception of interacting with a human-like *mind* rather than an embodied person in such task-oriented contexts.

Data availability

The data underlying this article and the supplementary materials are available at https://tinyurl.com/e78z95a8

Funding

Data collection of this research was funded by the Arthur W. Page Center for Integrity in Public Communication at the Donald P. Bellisario College of Communications at Pennsylvania State University, University Park.

Conflict of Interest

The authors do not have conflict of interest to disclose.

References

Araujo, T., Helberger, N., Kruikemeier, S., & De Vreese, C. H. (2020). In AI we trust? Perceptions about automated decision-making by artificial intelligence. *AI & Society*, 35(3), 611–623. https://doi.org/10.1007/s00146-019-00931-w

Berger, C. R. (1987). Communicating under uncertainty. In M. E. Roloff & G. R. Miller (Eds.), *Interpersonal processes: New directions in communication research* (pp. 39–62). Sage.

Berger, C. R. & Bradac, J. J. (1982). *Language and social knowledge: Uncertainty in interpersonal relations*. Edward Arnold.

Berger, C. R. & Calabrese, R. J. (1975). Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Human Communication Research*, 1(2), 99–112. https://doi.org/10.1111/j.1468-2958.1975.tb00258.x

Clatterbuck, G. W. (1979). Attributional confidence and uncertainty in initial interaction. *Human Communication Research*, 5(2), 147–157. https://doi.org/bzjzw9

Dennett, D. C. (1988). Précis of the intentional stance. *Behavioral and Brain Sciences*, 11(3), 495–505. https://doi.org/10.1017/S0140525X00058611

Dzindolet, M. T., Peterson, S. A., Pomranky, R. A., Pierce, L. G., & Beck, H. P. (2003). The role of trust in automation reliance. *International Journal of Human-Computer Studies*, 58(6), 697–718. https://doi.org/10.1016/S1071-5819(03)00038-7

Eagly, A. H., & Chaiken, S. (1984). Cognitive theories of persuasion. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 17, pp. 268–361). Academic Press.

Edwards, C., Edwards, A., Spence, P. R., & Westerman, D. (2016). Initial interaction expectations with robots: Testing the human-to-human interaction script. *Communication Studies*, 67(2), 227–238. https://doi.org/10.1080/10510974.2015.1121899

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, 114(4), 864–886. https://doi.org/dw8wt9

Fiske, S. T., & Neuberg, S. L. (1990). A continuum model of impression formation from category-based to individuating processes: Influences of information and motivation on attention and interpretation. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. 23, pp. 1–74). Academic Press.

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition* (1st ed.). Addison-Wesley.

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: a meta-analysis quantifying the effect of agency on social influence. *Human-Computer Interaction*, 30(5), 401–432. https://doi.org/gctmv2

Gudykunst, W. B. (1995). Anxiety/uncertainty management (AUM) theory: Current status. In R. L. Wiseman (Ed.), *Intercultural communication theory* (pp. 8–58). Sage.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford Publications.

Jian, J. Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, 4(1), 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Kiesler, S., & Goetz, J. (2002). Mental models of robotic assistants. In L. Terveen & D. Wixon (Eds.), *CHI'02 extended abstracts on human factors in computing systems* (pp. 576–577). Association for Computing Machinery. https://doi.org/10.1145/506443.506491

Kim, K. J., & Sundar, S. S. (2014). Does screen size matter for smartphones? Utilitarian and hedonic effects of screen size on smartphone adoption. *Cyberpsychology, Behavior, and Social Networking*, 17(7), 466–473. https://doi.org/10.1089/cyber.2013.0492

Langer, E. J., Blank, A., & Chanowitz, B. (1978). The mindlessness of ostensibly thoughtful action: The role of "placebic" information in interpersonal interaction. *Journal of Personality and Social Psychology*, 36(6), 635–642. https://doi.org/bbh438

Lee, J., & Moray, N. (1992). Trust, control strategies and allocation of function in human-machine systems. *Ergonomics*, 35(10), 1243–1270. https://doi.org/czxpwv

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, K. M. (2004). Presence explicated. *Communication Theory*, 14(1), 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lee, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of Communication*, 56(4), 754–772. https://doi.org/bt6qfs

Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 5(1), 1–16. https://doi.org/10.1177/2053951718756684

Lins de Holanda Coelho, G., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment*, 27(8), 1870–1885. https://doi.org/10.1177/1073191118793208

Liu, B., & Wei, L. (2021). Machine gaze in online behavioral targeting: The effects of algorithmic human likeness on social presence and social influence. *Computers in Human Behavior*, 124. https://doi.org/10.1016/j.chb.2021.106926

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, 2, 7–33. https://doi.org/10.30658/hmc.2.2

McCroskey, J. C., & Teven, J. J. (1999). Goodwill: A reexamination of the construct and its measurement. *Communication Monographs*, 66(1), 90–103. https://doi.org/b2dsh2

Miller, T., Howe, P., & Sonenberg, L. (2017). *Explainable AI: Beware of inmates running the asylum*. IJCAI 2017 Workshop on Explainable Artificial Intelligence (XAI). https://arxiv.org/abs/1712.00547

Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1–21. https://doi.org/10.1177/2053951716679679

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. https://doi.org/gctmrs

Oh, C. S., Bailenson, J. N., & Welch, G. F. (2018). A systematic review of social presence: Definition, antecedents, and implications. *Frontiers in Robotics and AI*, 5, Article 114. https://doi.org/10.3389/frobt.2018.00114

Ososky, S., Philips, E., Schuster, D., & Jentsch, F. (2013). A picture is worth a thousand mental models: Evaluating human understanding of robot teammates. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 57(1), 1298–1302. https://doi.org/10.1177/1541931213571287

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. Cambridge University Press. [Database]

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1–28. https://doi.org/10.1037/h0092976

Schlosser, M. (2019, October 28). *Agency*. Stanford Encyclopedia of Philosophy. https://rb.gy/qetv1d

Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Houghton, Mifflin and Company.

Shin, D., & Park, Y. J. (2019). Role of fairness, accountability, and transparency in algorithmic affordance. *Computers in Human Behavior*, 98, 277–284. https://doi.org/gkch

Smith, A. (2018, November 16). *Public attitudes toward computer algorithms.* Pew Research Center. https://rb.gy/zwmpuf

Sundar, S. S. (2004). Loyalty to computer terminals: Is it anthropomorphism or consistency?. *Behaviour & Information Technology*, 23(2), 107–118. https://doi.org/b9v6c4

Sundar, S. S., & Nass, C. (2000). Source orientation in human-computer interaction: Programmer, networker, or independent social actor. *Communication Research*, 27(6), 683–703. https://doi.org/10.1177/009365000027006001

Takayama, L. (2015). Telepresence and apparent agency in human–robot interaction. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 160–175). Wiley Blackwell.

Turing, A. M. (1950). *Computing machinery and intelligence*. Mind, 59(236), 433–460. https://doi.org/10.1007/978-1-4020-6710-5_3

United States Census Bureau (2018, April 19). *2013–2017 American community survey 5-year estimates*. https://rb.gy/cbenof

White, R. W. (1959). Motivation reconsidered: The concept of competence. *Psychological Review*, 66(5), 297–333. https://doi.org/10.1037/h0040934