Human Communication Research, 2023, **00**, 1–10
https://doi.org/10.1093/hcr/hqad034
Original Research

International Communication Association

OXFORD

Original Research

Minding the source: toward an integrative theory of human–machine communication

Eun-Ju Lee [1,*]

[1]Department of Communication & Center for Trustworthy AI, Seoul National University, Seoul 08826, South Korea
*Corresponding author: Eun-Ju Lee. Email: eunju0204@snu.ac.kr

Abstract

According to the computers are social actors (CASA) paradigm, a dominant theoretical framework for research on human–computer interaction, people treat computers as if they were people. Recent studies on human–machine communication (HMC) and human–artificial intelligence (AI) interaction, however, appear to focus on when and how people respond to machines *differently* than to human agents. To reconcile this apparent contradiction, this study reviews critically the two overarching theoretical explanations proposed and tested in each respective tradition, the mindlessness account and the machine heuristic. After elaborating on several conceptual and operational issues with each explanatory mechanism, an alternative theoretical model of HMC is proposed that integrates both research traditions and generates predictions that potentially deviate from the dual-process models. Lastly, it is discussed how recent developments in AI technology invite modifications to the current understanding of HMC and beyond.

Keywords: artificial intelligence, computers are social actors, human–machine communication, machine heuristic, mindlessness.

Once a sacred realm that belonged solely to human beings, communication is no longer an act that only humans can perform. Going beyond merely transferring messages between humans, machines can now serve as both mediators and communicators that actively participate in the construction of messages designed for either mass audience or an individual recipient (Sundar & Lee, 2022). If artificial intelligence (AI) refers to machines that are capable of performing tasks that require human intelligence (Turing, 1950), then the ability to communicate would certainly be among AI's key qualifications.

With the ever-increasing integration of AI into human communication processes, a burgeoning field of research has formed that addresses how people evaluate and respond to AI that serves in diverse communicative roles, such as a news reporter in automated journalism (Cloudy et al., 2021), a fact-checker that tags suspicious information (Banas et al., 2022), a friend one can banter with (Brandtzaeg et al., 2022), or a personal assistant who corrects misspelled words without even being asked (Hancock et al., 2020). However, before evaluating which agent, AI or human, is more effective in accomplishing specific communicative goals, one should first ask if people respond any differently to AI than to humans, and if so, how and why.

In fact, long before AI appeared on the horizon, Reeves and Nass (1996) reported a series of robust findings in their seminal book, "The media equation: How people treat computers, television, and new media like real people and places," that people treat computers as if they were real people. Despite their awareness that computers are not humans, people consistently apply a range of social rules and scripts, such as gender stereotypes (Nass et al., 1997), attraction based on similarity (Moon & Nass, 1996), politeness (Nass et al., 1999), and reciprocity (Fogg & Nass, 1997a). Collectively, these findings led to the conclusion that computers are "social actors" (CASA), if not full-blown humans, which elicit reactions normally expected in interhuman encounters.

Unlike earlier human–computer interaction (HCI) research, which is heavily grounded in the "equation" between humans and computers, recent works on human–AI interaction (HAII) seem to be gravitated toward demonstrating how people are *differently* oriented toward AI than they are toward humans, for better or worse. Considering that HAII is a natural extension of HCI, the apparent contradiction is rather intriguing, and even ironic. After all, AI as a more advanced form of technology can resemble humans more closely than its predecessors in both appearance and functionality, rendering its nonhumanness much less evident. To reconcile this inconsistency, the current essay first revisits the notions of mindlessness and machine heuristic, which are the two commonly invoked explanations for why people respond to computers and AI agents the way they do. Specifically, after contemplating some conceptual and empirical issues with the mindlessness account, I attempt a conceptual explication of machine heuristic in light of cognitive heuristics and dual-process models. Then, I propose an alternative model of human–machine communication (HMC), which integrates both CASA and machine heuristic research traditions. Lastly, it is briefly discussed how future developments in AI technology might demand modifications to the current understanding of HMC and beyond.

Mindless social responses to computers?

Arguably one of the most influential theoretical frameworks that guided HCI research for the past couple of decades, the

Received: 8 April 2023. **Revised:** 1 September 2023. **Accepted:** 8 September 2023
© The Author(s) 2023. Published by Oxford University Press on behalf of International Communication Association.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

CASA paradigm has established its robustness in numerous empirical investigations. Not as conclusive, however, is why people emit such seemingly irrational responses, treating lifeless machines as if they had gender, personalities, feelings, and intentions. Among several explanations proposed, mindlessness appears to have survived (Lee, 2008, 2010; Nass & Moon, 2000; Sundar & Nass, 2000). Just as people were more likely to comply with a request that contained "because" than the one without, regardless of how legitimate the following reason was (Langer et al., 1978), people mindlessly process and respond to incoming stimuli, thereby failing to adjust their default *social* reactions while dealing with an *asocial* being that evinces minimal human-like cues.

In this view, mindlessness represents a general state of mind, rather than a variable. Boldly put, "individuals' interactions with computers, television, and new media are *fundamentally social and natural*" (Reeves & Nass, 1996, p. 5), such that the tendency to equate mediated objects and experiences with real ones is expected to prevail, regardless of one's age, education level, or technological proficiency. However, as subsequent studies focus largely on whether or not people exhibit social responses in yet another context with a different social rule, the mindlessness account has not been subjected to rigorous scientific scrutiny, but rather assumed as a legitimate theoretical explanation. Several issues seem to deserve our attention.

Falsifiability of the mindlessness account

First, mindlessness as a psychological state was not separately measured, nor was it directly tested. Instead, mindlessness was either presumed or inferred from people's reactions to computers. That is, if people are mindful, they should not apply social rules to a computer—given that they did, they must have been mindless. It seems analogous to saying that if argument quality made a difference in persuasion outcomes, people must have engaged in systematic processing. However, to validate the causal connection between mindlessness and social responses to a computer, the most straightforward method would be to systematically vary the degree of mindlessness and see if it alters the extent to which people treat computers like humans in the predicted direction.

In one such attempt, Lee (2008) manipulated the number of tasks imposed on the participants (one vs. two; Experiment 1) and the modality of computer output (speech vs. text; Experiment 2) and examined if those cognitively busier with multitasking or less attentive to the computer output while processing text (vs. voice) became more mindless, and thus, more likely to exhibit gender-typed responses to computers. Although the results supported the mindlessness predictions, even if they had been otherwise, one cannot determine whether it was because the manipulated stimuli (i.e., the number of tasks, the output modality) failed to induce the state of mindlessness, or because mindlessness failed to trigger social treatment of computers. Unless we measure mindlessness independently and test these two processes separately, the presumed connection between mindlessness and social responses to computers remains virtually unfalsifiable.

Monolithic treatment of social responses

Second, it is not entirely clear if "social responses" can be treated as a single, monolithic entity. As an umbrella term, social responses refer to the reactions people naturally show toward other humans based on specific attributes (e.g., gender, personality, shared membership) or general norms (e.g., politeness, reciprocity), as documented in the communication or psychology literature. However, not all social responses are equally indicative of mindlessness. If one reported more positive affect, more positive evaluations of the interaction, and more positive regard for the computer when it randomly praised their performance (flattery), rather than provided generic feedback (Fogg & Nass, 1997b), such reactions might represent a mindless, "click-whirr" response (Cialdini, 2007) driven by the ego-enhancement motive. But what about *discounting* the flattering computer's output as *less* trustworthy?

In Lee's (2010) study, for instance, participants found the flattering (vs. nonflattering) computer to be more socially attractive, but were more suspicious of its claims and dismissed its suggestions (Experiment 2). Suspecting hidden motives of a flattering person requires a higher degree of mindfulness than taking the groundless praises at face value, for suspicion is "a dynamic state in which the individual entertains multiple, plausibly rival hypotheses about the motives or genuineness of a behavior" (Hilton et al., 1993, p. 502). Still, one can argue that suspecting a computer's motives overlooks the fact that computers have no intention, and hence signals mindlessness. If so, the only way one can demonstrate mindfulness is to treat flattering and nonflattering computers equally, but it is unclear if such indiscriminate reactions represent a more mindful act, compared with either favoring or disfavoring a computer based on its output. Quite contrarily, one may argue that flattery is a "content cue" that is heeded more carefully when people adopt a systematic processing strategy (Chaiken, 1980). In fact, the greater suspicion about and the subsequent dismissal of the flattering (vs. nonflattering) computer's suggestions dissipated when the participants were distracted by a secondary task, and hence more mindless (Lee, 2010); that is, flattery made a difference in HMC when people were *more* mindful, which counters the mindlessness account.

Underspecified psychological mechanism

Lastly, the potential variability in the *activation* of social scripts is not fully engaged. In the CASA framework, social responses were conceptualized as built-in, default reactions that are automatically activated upon encountering human-like cues (Nass & Moon, 2000). Mindfulness is thus required in the *application* stage to suppress, or "undo," the activated scripts that are deemed irrelevant or inappropriate to the dealings with a computer. After all, to assess whether activated knowledge is usable or not is a controlled process (Higgins, 1996) that demands additional cognitive efforts. However, some might not exhibit gender stereotypes to a computer, not because they have deliberately judged the gender-associated social script to be inappropriate and thus unusable in HCI, but because they did not subscribe to gender stereotypes in the first place, regardless of their interactant's ontological identity (i.e., no activation).

In sum, while a large volume of research has confirmed the key proposition of the CASA paradigm that concerns how people respond to computers, only few studies (e.g., Johnson et al., 2004; Lee, 2008, 2010; Xu et al., 2022) have attempted to validate the mindlessness explanation directly, precisely conceptualize what "social responses" are, and/or articulate the process through which mindlessness induces social responses. Nonetheless, recent works on HMC no longer endorse the blanket statement that "individuals' interactions

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

with computers . . .are *fundamentally social and natural*" (Reeves & Nass, 1996, p. 5). Instead, researchers have incorporated the machine heuristic (i.e., the extent to which people attribute machine-like qualities to computing systems) as either an explanatory mechanism or a contingent condition for proposed source effects (human vs. machine), or a lack thereof (e.g., Liu & Wei, 2018; Sundar & Kim, 2019; Wang, 2021).

Machine heuristic: conceptual explication and integration with CASA

Apart from some exceptions (e.g., Johnson et al., 2004; Lee & Nass, 2002; Morkes et al., 1999), most CASA studies did not directly compare human–human interaction with HCI, but instead aimed to replicate the rules of interpersonal interaction *within* the HCI context. That is, insofar as people exhibited the same pattern of responses to computers as they would to humans, like favoring the computer with similar than dissimilar "personality" (Moon & Nass, 1996), it was less of a concern whether the degree to which people evince such reactions varies depending on the source.

As AI comes to play increasingly diverse roles in communication processes in lieu of human actors, however, the pendulum has begun to swing in the opposite direction. Rather than confirming the "media equation," much HMC research focuses on how the source attribution (human vs. machine) elicits different cognitive, affective, and behavioral reactions (e.g., Banas et al., 2022; Cloudy et al., 2021; Jones-Jang & Park, 2023; Waddell, 2018). This apparent shift makes sense in light of the surprise value in scientific research. When computers were unmistakably different from humans, it was surprising and thus worth reporting that people still failed to treat them differently. Contrarily, when "AI agents are developed to be more intimate, proximate, embodied, and human-like" (Jones-Jang & Park, 2023, p. 1), it is surprising that people, nonetheless, differentiate between AI and human. In general, "empirical confirmation of surprising predictions constitutes a non-trivial advance" (Trafimow, 2013, p. 1).

Despite the diverging orientations, HMC studies still tend to assume that people mostly engage in heuristic (vs. systematic) processing when dealing with machines, a relatively mindless mode of processing, and examine what specific heuristics are invoked when people make sense of machine-generated messages. In so doing, the machine heuristic, which refers to generalized beliefs or "stereotypes about the operation of machines" (Sundar, 2020, p. 79), was proposed as an explanation for why people respond differently to an AI agent than its human counterpart. For instance, those with stronger beliefs in machine heuristic were more likely to entrust an AI (vs. human) agent with their personal information (Sundar & Kim, 2019), experienced lower levels of emotional involvement with the AI-authored (vs. human-authored) news article (Liu & Wei, 2018), and evaluated the news to be less biased and more credible when the uncivil comment section was moderated by a machine (vs. human) agent (Wang, 2021).

The introduction of machine heuristic as a user characteristic that systematically alters the processes and outcomes of HMC represents a refinement of the CASA paradigm. That is, while inheriting the key proposition that people are generally mindless when interacting with machines, and thus prone to rely on heuristics, recent studies demonstrated that not all people treat humans and computers equally, and more important, proposed the source of variation (machine heuristic) a priori for empirical validation. Still, several conceptual and empirical questions need to be addressed for machine heuristic to be well integrated with the CASA framework and inform an integrative theory of HMC.

Heuristic vs. personal beliefs vs. folk theory

As a key construct in the heuristic–systematic model (HSM; Chaiken, 1980), heuristics refer to "relatively general rules (scripts, schemata) developed by individuals through their past experiences and observations" (p. 753). They represent rules of thumb or mental shortcuts that help reduce cognitive load associated with information processing and decision making (Metzger et al., 2010). An implicit assumption is that heuristics are shared widely and available to most people, such that they are readily activated when people are unable or unwilling (or both) to process incoming stimuli thoroughly. As such, if there is considerable variability in how strongly people embrace and use the machine heuristic in their evaluations and judgments, then it might be more appropriate to define it as *personal* beliefs about machines.

Also, if machine heuristic represents lay people's understanding or expectations of how machines perform and operate with what consequences, then it sounds virtually indistinguishable from "folk theories." Defined as "intuitive, informal theories that individuals develop to explain the outcomes, effects, or consequences of technological systems" (DeVito et al., 2017, p. 3165), folk theories are believed to affect how individuals use the systems and respond to their outputs (Huang et al., 2022). If machine heuristic is measured by the degree to which people accept such statements as "As machines are precise, their prediction will be more reliable than humans'," and "As machines are unbiased, machines' guide will be more trustworthy" (S. Lee et al., 2023), which capture not just individuals' beliefs about key traits of machines, but also those about causal consequences inferred from them, then it becomes indistinguishable from folk theories.

What does "machine-like" mean?

As simple schemas or decision rules, heuristics encapsulate evaluative principles, such as "(message) length implies strength," "more arguments are better arguments," and "if other people think the message is correct then it is probably valid" (Chaiken, 1987, p. 4). Although variations exist, machine heuristic is often operationalized by how strongly people associate with machines some "machine-like" qualities, such as being objective, unbiased, accurate, error-free, unemotional, unyielding, and reliable (e.g., Banks et al., 2021; Cloudy et al., 2021; Sundar, 2020; Waddell, 2018; Wang, 2021). What merits note is that some studies (e.g., S. Lee et al., 2023; Sundar & Kim, 2019) operationalized the machine heuristic to include an explicit comparison with humans (e.g., "Machines are more trustworthy than humans," "When machines perform a task, the results are more objective than when humans perform the same task"). If machine heuristic hinges on what separates machines from humans, then stronger beliefs in machine heuristic will inhibit, rather than facilitate, social responses to computers, directly challenging the mindless explanation for the CASA paradigm.

Moreover, some researchers included specific behavioral outcomes that would reasonably follow certain beliefs about machines as part of the operational definition of machine

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

heuristic (e.g., "Machines can handle information in a secure manner, therefore it is generally okay to disclose my private information [e.g., card number, address, phone number, etc.] to them."; S. Lee et al., 2023). Apart from its double-barreled nature, if people indeed share more private information with the AI agent (behavioral outcome) because they believe "it is generally okay to disclose their private information to machines," it seems dangerously close to a tautology. If machine heuristic were to serve as an explanation for why people behave toward computers and humans differently, then it would be more appropriate to separate (a) individuals' beliefs about inherent characteristics of machines and (b) specific actions that stem from such beliefs, and define machine heuristic strictly in terms of the former.

Lastly, if machine heuristic refers to generalized beliefs or "stereotypes" about what machines are like and how they perform (Molina & Sundar, 2022; Sundar & Kim, 2019), its specific content might need to be updated, as people's expectations of machines evolve with technological advancements over time. For instance, if friendly and personable social chatbots that convey unwavering support and empathy become widely popular, then the currently commonplace notion that machines are "unyielding," "unemotional," and "cold" (Molina & Sundar, 2022; Sundar, 2020) may lose its ground. Likewise, when the generative AI's ability to create hallucination by providing utterly fabricated information is repeatedly highlighted in the media coverage (Weise & Metz, 2023), people might realize that the apparent objectivity of algorithms is only skin-deep and AI is not as accurate or reliable as they once believed. In both cases, the very definition of "machine-likeness" would need to be revised, but it is unclear if the current conceptualization of machine heuristic allows such updates. If it does, can we still call the machine heuristic a heuristic, when the if–then relationship it postulates (e.g., if a machine wrote the news article, it is unbiased) is in flux?

Does the machine heuristic operate heuristically?

To confirm that the machine heuristic indeed guides people's reactions to machines as a mental shortcut, it should be empirically demonstrated that the effects of machine heuristic on subsequent judgments and behaviors are amplified when people engage in heuristic rather than systematic processing. Put differently, machine heuristic should become more influential when people lack either cognitive resources (e.g., due to multitasking, distraction, insufficient prior knowledge) or the motivation to process the communication systematically (e.g., due to low personal relevance or interest), or both (Chaiken et al., 1989). However, heuristic processing was often taken for granted as a default mode of information processing (see the MAIN model, Sundar, 2008; Sundar et al., 2019), and the very fact that beliefs in the machine heuristic served as a significant moderating variable of source effects was interpreted as evidence for heuristic processing (e.g., Banas et al., 2022; Banks et al., 2021; S. Lee et al., 2023; Sundar & Kim, 2019).

If people rely more on the machine heuristic while they engage in heuristic (vs. systematic) processing, then it follows that the more mindless, the more likely people would treat humans and computers differently—a prediction that directly contradicts the mindlessness explanation for CASA. Alternatively, if the effects of machine heuristic on individuals' responses to machines are amplified when people are *more* mindful, then we should probably reconsider its status as a heuristic, in light of the dual-process models. Although HSM acknowledges that both heuristic and systematic processing occur concurrently (Chaiken et al., 1989), that does not explain why machine heuristic exerts *greater* influence on communication outcomes in the systematic, rather than heuristic, processing mode. Moreover, if peripheral and central processing represent the low and high ends of the elaboration continuum, as per the elaboration likelihood model (ELM; Petty & Cacioppo, 1986), the influence of machine heuristic should be *lessened* as people process information more effortfully (Petty et al., 1981). As such, an empirical test is in order that examines in which processing mode people rely more heavily on machine heuristic, and thus differentiate between machines and humans.

Machine heuristic as mediator vs. moderator

Another remaining question concerns the theoretical role that the machine heuristic plays in predicting and explaining people's responses to machines (vs. humans) (see Bellur & Sundar, 2014, for a thorough discussion on "heuristics-as-variables," p. 119). Similar to other heuristics, machine heuristic is considered to be stored in individuals' memory and "cues on the media interface suggesting a machine source will trigger this heuristic, which in turn will shape perceived quality and credibility of media content as well as the entire user experience" (Sundar, 2020, p. 80). Empirical investigations, however, seem to diverge in how such processes are modeled and tested.

On the one hand, some studies (e.g., Cloudy et al., 2021; Waddell, 2018, Study 2) proposed that machine heuristic *mediates* the effects of source (human vs. machine; IV) on the users' cognitive, affective, and behavioral reactions (DVs; Figure 1A). Two conceptual issues merit note. First, it is one thing that the exposure to machine agency cues *activates* the machine heuristic, and it is another that the source cue *affects* how strongly one endorses the machine heuristic, as a mediation model implies. Put differently, there is no inherent reason why people should become more (or less) receptive to the machine heuristic, simply because they encountered a machine agency cue, when the sheer act of asking them about machine heuristic would likely activate their beliefs about machines in the human condition as well. Instead, the degree to which people agree that machines are fair, objective, and error-free is likely to vary as a function of the machine agent's preceding performance. After reading a well-written news article ostensibly authored by AI, for instance, people would accept the machine heuristic more readily (*à la* availability heuristic; Tversky & Kahenman, 1973), as compared with those who read a human-authored article. To rule out this potential confounds, one should examine if exposure to a machine agent's

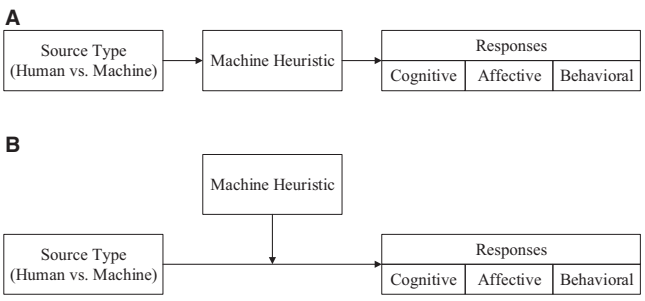


Figure 1. (A) Machine heuristic as mediator of source effect and (B) machine heuristic as moderator of source effect.

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

poor performance still induces a stronger endorsement of the machine heuristic than does its human counterpart's.

Second, it is unclear why beliefs in machine heuristic (mediator) should affect the evaluative outcomes (DVs) in both human and machine conditions alike. By definition, the machine heuristic captures people's expectations and beliefs about machines. If so, beliefs in machine heuristic should be either (a) irrelevant to the evaluations of a human agent or (b) inversely associated with them, if people automatically compare machines with humans (see Lee et al., 2022 for how a human agent's unsatisfactory performance enhances individuals' ratings of AI). Either way, the association between the mediator (machine heuristic) and the DVs should vary depending on the agency type (machine vs. human), but the simple mediation model fails to uncover this interaction. In fact, studies found that beliefs in machine heuristic significantly altered the evaluations of an AI agent's performance, but not a human's (e.g., Lee et al., 2023b; Sundar & Kim, 2019).

On the other hand, others highlighted the variability in people's endorsement of machine heuristic as a contingent condition for the source effect to occur. In this scenario, the machine heuristic *moderates* the degree to which users differentially respond to machine and human agents (Figure 1B). Although it was not framed this way, the moderating effect of machine heuristic can be deemed as a special case of a confirmation bias—that is, those who believe that AI in general is more objective, accurate, and reliable would think more highly of and respond more favorably to a specific AI agent's performance, be it AI journalist (Lee et al., 2023b), AI content moderator (Wang, 2021), or AI travel agent (Sundar & Kim, 2019). In contrast, those who do not espouse such beliefs would either treat machines and humans similarly or even evaluate machines more negatively than their human counterparts.

When do people mind the source?

As discussed above, recent studies on HMC that utilize the machine heuristic inherit the key proposition of the CASA paradigm by assuming heuristic (mindless) processing as a default mode, but they also depart from it by challenging the universality of social responses to machines. More important, machine heuristic presupposes, either explicitly or implicitly, that people associate with machines some unique attributes that distinguish machines from humans. Such conceptualization leads to the prediction that the more mindless, the more likely people rely on the machine heuristic, thereby exhibiting greater differentiation between human and machine agents. This directly counters the mindlessness account of CASA.

Despite such differences, both traditions assume that people are constantly mindful of the source. It is how people treat different sources (human vs. machine), either similarly or differently, that these accounts diverge. However, the extent to which people mind the source, whether it is a human or a machine, might vary as a product of situational and dispositional factors, and more important, in a way that deviates from what classical dual-process theories would predict.

Self-confirmation as super-heuristic

Perhaps better known as confirmation bias, Metzger and Flanagin (2013) refer to the "tendency for people to view information as credible if it confirms their preexisting beliefs and not credible if it counters their existing beliefs" as "self-confirmation heuristic" (p. 215). According to the authenticity model of computer-mediated communication (Lee, 2020), authenticity of communication consists of three subcomponents: how well the claimed identity matches the source's real identity (authenticity of source, p. 61), how truthfully a message represents its object (authenticity of message, p. 62), and how closely people feel they are a part of actual interaction (authenticity of interaction, p. 63). In the authentication process, although the model does not use the term, self-confirmation heuristic can serve as a super-heuristic that precedes and guides the operation of other heuristics, if ever. Specifically, "when an incoming message deviates from individuals' expectancy, which is formed based on general knowledge, schemas, social scripts, stereotypes, and their past experiences," (p. 64) they might simply dismiss the communication as false or inauthentic, and the authentication process ends immediately. As the sufficiency principle of HSM postulates (Chaiken et al., 1989), however, under some circumstances where the accuracy motivation is high, people are willing to take further steps, "by utilizing available cues and engaging in different cognitive processes" (p. 64).

Unlike HSM and ELM, which posit that people rely more on heuristics and peripheral cues when they lack the motivation and/or the ability to process messages thoroughly, however, the authenticity model proposes that it is when individuals are *more* willing to expend extra cognitive efforts to authenticate the communication that they take various cues and associated heuristics into account. In fact, while some dual-process studies differentiated message (e.g., argument quality) and nonmessage cues (e.g., expertise and likability of source) (e.g., Chaiken, 1980; Petty et al., 1981; see Petty et al., 1999 for how this content-based partition should not be equated with the conceptual distinction between arguments and cues), the authenticity model refers to both message and nonmessage features as "authenticity markers" (Lee, 2020, p. 69; also see Kruglanski & Thompson, 1999) that people may turn to when their expectancy has been violated. In this view, while "counting on heuristic cues (e.g., the number of likes) while ignoring more substantial information (e.g., argument quality) may signal cursory information processing," "if only heuristic cues are available . . . utilizing such cues may reflect a more effortful cognitive process than turning a blind eye to them" (p. 65).

Recent empirical evidence seems to support this prediction. For instance, Lee et al. (2023a) found that communication channel (a nonmessage cue) exerted greater influence on individuals' acceptance of a persuasive health message among those *more* interested in health. Specifically, participants inferred higher levels of ulterior motives (i.e., less authentic) from a medical reporter's newspaper column than from his Facebook post with identical content, which lowered their intention to follow the recommended behavior. But such a tendency was stronger among those *more* interested in health, who presumably processed the health messages more carefully.

Directly germane to HMC, studies have also reported the significant effects of source type (human vs. AI) when individuals are *more* motivated to process information systematically. For example, after reading a news article attributed to either algorithm or a human journalist, participants were more likely to rate the article credibility differently depending on the source, when it lacked objectivity (i.e., no source attribution, value-laden words like "thankfully"), which would have triggered message scrutiny (Tandoc et al, 2020).

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

Similarly, Lee et al. (2022) examined how people evaluate the AI (vs. human) moderator of user comments sections and found that participants were more suspicious about the AI moderator's ulterior motives, but only when the remaining comments were counter-attitudinal or when no explanation was provided for deleted comments. When the remaining comments were mostly proattitudinal or when reasons were provided for why some comments had been removed, participants' suspicion of ulterior motive remained unaltered, whether the moderator was AI or a human.

An integrative theory of HMC: machine heuristic as conditional moderator

Taken together, the assumption that people mind who or what the source is, even if they fail to adjust their responses accordingly, might need to be revisited. People might not care, unless they have a reason to consider who authored a news article, who checked the facts, who composed a music, and the like. When the message violates their expectations, and thus motivates them to mind the source among other factors, then their beliefs about the source, either categorical (e.g., machine, AI) or individuated (e.g., ChatGPT, Siri), will set in and guide the subsequent judgments. Such beliefs are likely to form based on one's personal experiences with machines, observations of others' experiences, and media portrayals of machines, to name a few. What is more, in order for such beliefs to shape evaluative outcomes, they should be relevant to the task at hand. For instance, one's beliefs about AI's fairness might not affect how people evaluate a social chatbot's friendliness.

There are two opposite ways in which one's prior beliefs about machines bias their reactions to a machine's performance that has violated their expectations: assimilation and contrast. First, people might evaluate the machine agent's output in a belief-congruent manner, such that those holding more positive beliefs about how machines operate (e.g., being fair, objective, accurate, and unbiased) might be *more* receptive to the machine agent's expectancy-disconfirming act, when compared with those holding no such beliefs. Albeit not limited to the cases wherein expectancy violations occurred, studies reported that those with stronger beliefs in machine heuristic (i.e., machines are objective, unbiased, accurate, error-free) were more likely to accept an AI fact-checker's truth verdicts (Banas et al., 2022) and rated a news article to be less biased when the uncivil comment section was moderated by a machine (vs. human) agent (Wang, 2021). Alternatively, one's prior beliefs about machines might serve as an anchor against which the current event is evaluated. In such a case, those holding a positive machine heuristic will respond *more* negatively to an AI agent's underperformance, which fell short of their inflated expectations. The finding that a news article lacking objectivity was penalized more severely with lower credibility ratings when attributed to algorithm than to a human journalist (Tandoc et al., 2020) might suggest a negative expectancy violation, if we can assume that participants overall believed that machines are objective, balanced, and accurate. Figure 2 summarizes these processes.

Take an AI fact-checker, for example. When its verdict confirms an individual's existing beliefs, attitudes, and/or opinions, they will accept it as is, without considering who rendered the verdict—AI or human (i.e., no main effect of the source type). When the verdict counters their beliefs, attitudes, and/or opinions, however, they would consider additional factors, like the source, to make sense of the unexpected outcome. With an AI fact-checker, their prior beliefs about AI then guide how they respond to the expectancy-disconfirming message. If people assimilate their responses to their existing stereotypes about AI, then those holding more positive stereotypes will be more likely to accept the disconfirming fact-check verdict. Consequently, the confirmation bias in the acceptance of corrective information will be attenuated in the AI, rather than human, condition (Figure 3A). In contrast, those with more positive beliefs about AI might exhibit even more negative reactions to the AI agent's opinion-challenging verdict due to a negative expectancy violation, thereby amplifying the confirmation bias in the AI condition (Figure 3B).

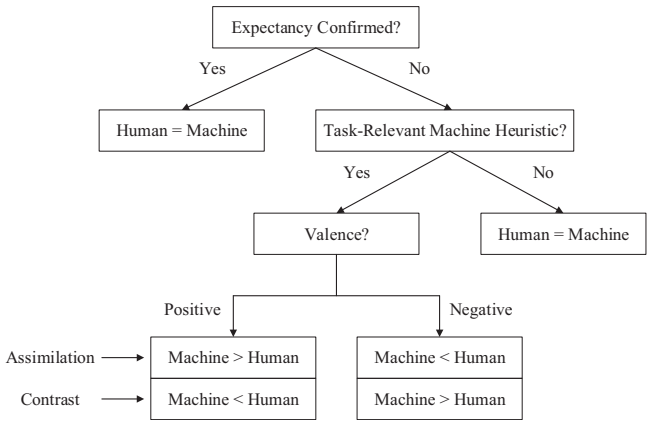


Figure 2. Machine heuristic as conditional moderator of source effect.

The current model calls for three empirical considerations. First, we should rethink what a failed manipulation of source actually means. When study participants do not recall whether the message was written by algorithm or a person, their data are usually discarded (e.g., Tandoc et al., 2020). However, when a sizeable proportion of participants turned out to have paid little attention to the source, we should perhaps ask how mindful people normally are of the message source in real life, instead of artificially exaggerating the source label so it will rarely go unnoticed. Second, it is important to distinguish between (a) relying solely on the source cue (AI vs. human) with disregard for the message content and (b) relying on both message and source cues, for they each represent qualitatively different cognitive processes. By establishing that there is no difference in message elaboration, one can conclude that the greater source effect is not a sign of heuristic or mindless processing, but a product of *more* effortful processing that took all the available cues into consideration. Third, it remains to be tested in which direction prior beliefs about machines bias people's reactions to negative expectancy violations by AI agents, assimilation or contrast, and moreover, if it varies, what factors account for such variation. For instance, when one's beliefs about AI are firmly grounded in their direct experiences, as opposed to remote observations or hearsay (i.e., belief certainty), assimilation might be more likely to occur.

At the same time, some limitations should be noted regarding the proposed model's generalizability. First, not all communications may allow the binary classification of either confirming or disconfirming expectancy. People might not even have any prior knowledge or experience to form solid

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

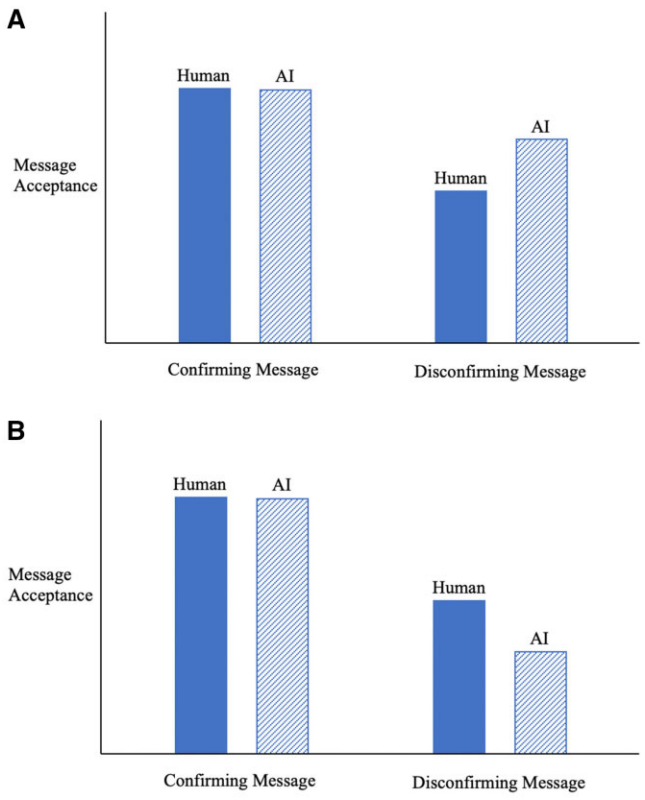


Figure 3. (A) Assimilation effect of positive machine heuristic and (B) contrast effect of positive machine heuristic.

expectations about the subject of a message. In such cases, factors known to facilitate message elaboration, such as the accuracy motivation or personal relevance, are likely to determine on which route their message processing will take place.

Second, the current model is better suited for communication contexts where the source salience is relatively low. Even though people are told who/what the source is, reading a byline of a news article ("authored by algorithm") is not the same as interacting with a chatbot in terms of how vividly people feel the machine's presence. As such, the present model might be more useful in explaining one-to-many, rather than one-to-one, communication contexts (i.e., AI Curator and AI Creator; Sundar & Lee, 2022). When the agent type is sufficiently salient to begin with, individuals' beliefs about AI will set in instantly, thereby shaping communication processes and outcomes. For instance, the agency type might influence the degree of message elaboration, such that people might process a message more effortfully when it is attributed to AI than a human, due to the novelty. If so, content cues will exert stronger influence when AI has authored the message.

Looking ahead and moving forward

To develop a theory is to answer the fundamental "why" question about observed regularities in the focal phenomenon of interest (Berger et al., 2010). As such, to theorize HMC, one needs to identify consistent patterns in people's reactions to machines, propose the mechanisms to explain why such patterns exist, and put them to empirical tests for validation. Can mindlessness and/or machine heuristic serve as the core mechanism that predicts and explains the processes and effects of HMC? Are there any alternative explanations? More important, do we need a theory exclusively about HMC or can it be subsumed within the purview of a general communication theory?

In attempts to address these questions, the current essay reviewed how mindlessness and machine heuristic have been conceptualized and operationalized in the literature and raised some questions about their viability as a theoretical explanation for CASA in particular, and for HMC in general. After identifying potential incompatibilities, an alternative model informed by the authenticity model of computer-mediated communication (Lee, 2020) was proposed to reconcile the two research traditions and develop an integrative theory of HMC. Although the authenticity model was developed for human–human communication *via* computer, it may also help explain human communication *with* computer by integrating the source type (human vs. machine) as a variable.

With the seamless integration of AI technology into numerous applications and services, it has become critical to ensure transparency concerning the specific roles and functions of AI from the ethical standpoint, but less clear is how users process such information, and with what consequences. Both CASA and machine heuristic studies assume that people are mindful of the source, even if they often fail to adjust their expectations and reactions accordingly. However, recent findings (e.g., Lee et al., 2022; Tandoc et al., 2020) suggest that people may not be sufficiently mindful of the source, unless they need to. Insofar as they enjoy the conversation, obtain useful information from the news article, and appreciate the song they listen to, why should they care who does the talking, wrote the news, and composed the song? After all, we are cognitive misers (Fiske & Taylor, 1984). If so, minding the source, as a cue, and relying on associated cognitive shortcut (i.e., machine heuristic) to make judgments and decisions, which the ELM has considered as an indication of less mindful processing, may in fact reflect rather effortful and elaborate processing of communication.

Moreover, recent technological breakthroughs seem to challenge the CASA's key premise that it is unreasonable and irrational to treat computers like humans. For instance, we are witnessing an increasing number of virtual humans who look like a human, dance like a human, and talk like a human (https://www.virtualhumans.org/). Why then should following a virtual influencer be different from admiring BTS or BLACKPINK? When conversing with a social chatbot who is always there whenever I need him/her and listens to whatever I want to vent without worrying about a leak (Brandtzaeg et al., 2022), disregarding the agent's ontological identity can be a fairly rational choice to gratify one's hedonistic needs, maximizing the pleasure and emotional support one can derive from the "friendship." In the age of multiple identities that are in constant flux, it may not be too different from engaging with a role-playing partner in the metaverse, wherein users are fully aware that the avatars' appearance and personality may not match the "real" person behind the cartoon character, and yet, suspend their disbelief to get fully immersed in the virtual world and enjoy the experience.

Once people mind the source, then their beliefs about machines in general will direct how the source type affects their subsequent reactions. Thus far, studies employing the machine heuristic tend to focus on machines' agentic, rather than communal, qualities of largely positive connotations, such as being objective, accurate, precise, and error-free (see Molina & Sundar, 2022 for an exception). While those are

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

undoubtedly core traits people seem to associate with machines at the moment, more systematic approaches are called for, if we are to use this construct to account for a wide variety of HMC beyond task-oriented domains. Perhaps a more comprehensive set of adjectives people use to describe machines can be gleaned inductively from a large-scale survey with a representative sample, which will help identify dimensions on which people evaluate machines, either similarly or differently, vis-à-vis humans. In this regard, Hong et al. (2022) proposed "creative machine heuristics" to capture what human-like traits people expect of machines, focusing on creativity in particular. Although the attempt to expand the repertoire of machine heuristic seems well justified, they focused on "how people think about machines being creative" (p. 2) rather than how creative people think machines are (or can be).

If machine heuristic serves as a moderator of the source effects, something one brings into the communication context, then what would contribute to such beliefs? Usual suspects include the exposure to media discourse, education, computer proficiency, and direct or indirect experiences with machines. Less obvious is the role of dissatisfaction with a human equivalent. In one study (Lee et al., 2022), for instance, after witnessing a human moderator's unsatisfactory performance, participants were more likely to endorse AI heuristic, suggesting that people consider AI as a potential replacement of suboptimal human agents. Possibly, disappointment with human performance may lead people to project more optimistic or wishful images unto AI.

In celebration of *Human Communication Research*'s 50th anniversary, this special issue is dedicated to innovative theory development. Since the pioneering book by Reeves and Nass (1996) was published, numerous studies have tested the CASA paradigm in a wide range of contexts, lending support to its key proposition. Still, replicating a particular phenomenon (in this case, social responses to computers) across contexts, albeit attesting to its regularity, falls short of explaining why the observed phenomenon occurs. If "identified regularities need to be explained by recourse to mechanisms that account for the regularity in question" (Berger et al., 2010, p. 10) for theory construction, then the psychological processes proposed herein may serve as one such mechanism, if validated through rigorous empirical testing. In this sense, the current essay represents a long over-due attempt to update the CASA framework that paved the way for HMC research almost three decades ago. Not only does the extent to which people adjust their responses to machines in consideration of their nonhuman nature vary as a function of mindlessness, but the degree to which people consider the nonhumanness as a relevant factor in their communication may also vary, depending on how mindful they are (or choose to be).

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2022R1A5A708390812).

Conflicts of interest: No potential conflict of interest was reported by the author.

Data availability

No data are linked to the current article.

References

Banas, J. A., Palomares, N. A., Richards, A. S., Keating, D. M., Joyce, N., & Rains, S. A. (2022). When machine and bandwagon heuristics compete: Understanding users' response to conflicting AI and crowdsourced fact-checking. *Human Communication Research*, *48*(3), 430–461. https://doi.org/10.1093/hcr/hqac010

Banks, J., Edwards, A. P., & Westerman, D. (2021). The space between: Nature and machine heuristics in evaluations of organisms, cyborgs, and robots. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 324–331. https://doi.org/10.1089/cyber.2020.0165

Bellur, S., & Sundar, S. S. (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

Berger, C. R., Roloff, M. E., & Roskos-Ewoldsen, D. R. (2010). What is communication science? In C. R. Berger, M. E. Roloff, & D. R. Roskos-Ewoldsen (Eds.), *The handbook of communication science* (pp. 3–20). Sage.

Brandtzaeg, P. B., Skjuve, M., &Følstad, A. (2022). My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, *48*(3), 404–429. https://doi.org/10.1093/hcr/hqac008

Chaiken, S. (1980). Heuristic versus systematic information processing and the use of source versus message cues in persuasion. *Journal of Personality and Social Psychology*, *39*(5), 752–766. https://doi.org/10.1037/0022-3514.39.5.752

Chaiken, S. (1987). The heuristic model of persuasion. In M. P. Zanna, J. M. Olson, & C. P. Herman (Eds.), *Social influence: The Ontario symposium* (Vol. *5*, pp. 3–39). Lawrence Erlbaum Associates.

Chaiken, S., Lieberman, A., & Eagly, A. H. (1989). Heuristic and systematic information processing within and beyond the persuasion context. In J. S. Uleman & J. A. Bargh (Eds.), *Unintended thought* (pp. 212–252). Guilford Press.

Cialdini, R. B. (2007). *Influence: The psychology of persuasion*. Collins.

Cloudy, J., Banks, J., & Bowman, N. D. (2021). The str(AI)ght scoop: Artificial intelligence cues reduce perceptions of hostile media bias. *Digital Journalism*. https://doi.org/10.1080/21670811.2021.1969974

DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). "Algorithms ruin everything": #RIPTwitter, folk theories, and resistance to algorithmic change in social media. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, USA (pp. 3163–3174). https://doi.org/10.1145/3025453.3025659

Fiske, S. T., & Taylor, S. E. (1984). *Social cognition*. McGraw-Hill.

Fogg, B. J., & Nass, C. (1997a). How users reciprocate to computers: An experiment that demonstrates behavior change. In *CHI '97 extended abstracts on human factors in computing systems* (pp. 331–332), USA. https://doi.org/10.1145/1120212.1120419

Fogg, B. J., & Nass, C. (1997b). Silicon sycophants: The effects of computers that flatter. *International Journal of Human–Computer Studies*, *46*(5), 551–561. https://doi.org/10.1006/ijhc.1996.0104

Hancock, J. T., Naaman, M., & Levy, K. (2020). AI-mediated communication: Definition, research agenda, and ethical considerations. *Journal of Computer-Mediated Communication*, *25*(1), 89–100. https://doi.org/10.1093/jcmc/zmz022

Higgins, E. T. (1996). Knowledge activation: Accessibility, applicability, and salience. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social psychology: Handbook of basic principles* (pp. 133–168). Guilford Press.

Hilton, J. L., Fein, S., & Miller, D. T. (1993). Suspicion and dispositional inference. *Personality and Social Psychology Bulletin*, *19*(5), 501–512. https://doi.org/10.1177/0146167293195003

Hong, J.-W., Fischer, K., Ha, Y., & Zeng, Y. (2022). Human, I wrote a song for you: An experiment testing the influence of machines'

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

attributes on the AI-composed music evaluation. *Computers in Human Behavior*, *131*, 107239. https://doi.org/10.1016/j.chb.2022.107239

Huang, S. A., Hancock, J., & Tong, S. T. (2022). Folk theories of online dating: Exploring people's beliefs about the online dating process and online dating algorithms. *Social Media + Society*, *8*(2), 1–12. https://doi.org/10.1177/20563051221089561

Johnson, D., Gardner, J., & Wiles, J. (2004). Experience as a moderator of the media equation: The impact of flattery and praise. *International Journal of Human–Computer Studies*, *61*(3), 237–258. https://doi.org/10.1016/j.ijhcs.2003.12.008

Jones-Jang, S. M., & Park, Y. J. (2023). How do people react to AI failure? Automation bias, algorithmic aversion, and perceived controllability. *Journal of Computer-Mediated Communication*, *28*(1), 1–8. https://doi.org/10.1093/jcmc/zmac029

Kruglanski, A. W., & Thompson, E. P. (1999). Persuasion by a single route: A view from the unimodel. *Psychological Inquiry*, *10*(2), 83–109. https://doi.org/10.1207/s15327965pl100201

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/10.1037/0022-3514.36.6.635

Lee, E.-J. (2008). Gender stereotyping of computers: Resource depletion or reduced attention? *Journal of Communication*, *58*(2), 301–320. https://doi.org/10.1111/j.1460-2466.2008.00386.x

Lee, E.-J. (2010). What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations. *Communication Research*, *37*(2), 191–214. https://doi.org/10.1177/0093650209356389

Lee, E.-J. (2020). Authenticity model of (mass-oriented) computer-mediated communication: Conceptual explorations and testable propositions. *Journal of Computer-Mediated Communication*, *25*(1), 60–73. https://doi.org/10.1093/jcmc/zmz025

Lee, E.-J., Kim, H. S., & Joo, M. H. (2023a). Social media vs. mass media: Mitigating the suspicion of ulterior motives in public health communication. *Health Communication*, *38*(11), 2450–2460. https://doi.org/10.1080/10410236.2022.2074781

Lee, E.-J., Kim, H. S., Suh, Y. J., & Park, J. W. (2022). Something's fishy about it: How opinion congeniality and explainability affect motivated attribution to artificial intelligence versus human comment moderators. *Cyberpsychology, Behavior, and Social Networking*, *25*(8), 496–503. https://doi.org/10.1089/cyber.2021.0347

Lee, E.-J., Kim, H. S., Suh, Y. J., & Park, J. (2023b, May 25–29). Not so automatic perception of hostility in automated journalism: The role of source congeniality and belief in AI heuristic. In *The 73rd annual conference of international communication association*, Toronto, Canada.

Lee, E.-J., & Nass, C. (2002). Experimental tests of normative group influence and representation effects in computer-mediated communication: When interacting via computers differs from interacting with computers. *Human Communication Research*, *28*(3), 349–381. https://doi.org/10.1111/j.1468-2958.2002.tb00812.x

Lee, S., Oh, J., & Moon, W.-K. (2023). Adopting voice assistants in online shopping: Examining the role of social presence, performance risk, and machine heuristic. *International Journal of Human–Computer Interaction*, *39*(14), 2978–2992. https://doi.org/10.1080/10447318.2022.2089813

Liu, B., & Wei, L. (2018). Reading machine-written news: Effect of machine heuristic and novelty on hostile media perception. In M. Kurosu (Ed.), *Human–computer interaction. Theories, methods, and human issues* (pp. 307–324). Springer.

Metzger, M. J., Flanagin, A. J., & Medders, R. B. (2010). Social and heuristic approaches to credibility evaluation online. *Journal of Communication*, *60*(3), 413–439. https://doi.org/10.1111/j.1460-2466.2010.01488.x

Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*(Part B), 210–220. https://doi.org/10.1016/j.pragma.2013.07.012

Molina, M. D., & Sundar, S. S. (2022). Does distrust in humans predict greater trust in AI? Role of individual differences in user responses to content moderation. *New Media & Society*. https://doi.org/10.1177/14614448221103534

Moon, Y., & Nass, C. (1996). How "Real" are computer personalities? Psychological responses to personality types in human–computer interaction. *Communication Research*, *23*(6), 651–674. https://doi.org/10.1177/009365096023006002

Morkes, J., Kernal, H. K., & Nass, C. (1999). Effects of humor in task-oriented human–computer interaction and computer-mediated communication: A direct test of SRCT theory. *Human–Computer Interaction*, *14*(4), 395–435. https://doi.org/10.1207/S15327051HCI1404_2

Nass, C., Moon, Y., & Green, N. (1997). Are machines gender neutral? Gender-stereotypic responses to computers with voices. *Journal of Applied Social Psychology*, *27*(10), 864–876. https://doi.org/10.1111/j.1559-1816.1997.tb00275.x

Nass, C., Moon, Y., & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology*, *29*(5), 1093–1109. https://doi.org/10.1111/j.1559-1816.1999.tb00142.x

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Petty, R. E., & Cacioppo, J. T. (1986). *Communication and persuasion: Central and peripheral routes to attitude change*. Springer.

Petty, R. E., Cacioppo, J. T., & Goldman, R. (1981). Personal involvement as a determinant of argument-based persuasion. *Journal of Personality and Social Psychology*, *41*(5), 847–855. https://doi.org/10.1037/0022-3514.41.5.847

Petty, R. E., Wheeler, S. C., & Bizer, G. Y. (1999). Is there one persuasion process or more? Lumping versus splitting in attitude change theories. *Psychological Inquiry*, *10*(2), 156–163. https://doi/org/10.1207/S15327965PL100211

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

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

Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human–AI interaction (HAII). *Journal of Computer-Mediated Communication*, *25*(1), 74–88. https://doi.org/10.1093/jcmc/zmz026

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–9), UK. https://doi.org/10.1145/3290605.3300768

Sundar, S. S., & Lee, E.-J. (2022). Rethinking communication in the era of artificial intelligence. *Human Communication Research*, *48*(3), 379–385. https://doi.org/10.1093/hcr/hqac014

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

Sundar, S. S., Xu, Q., & Dou, X. (2019). Role of technology in online persuasion: A MAIN model perspective. In S. Rodgers & E. Thorson (Eds.), *Advertising theory* (2nd ed., pp. 70–88). Routledge.

Tandoc, E. C., Jr., Yao, L. J., & Wu, S. (2020). Man vs. machine? The impact of algorithm authorship on news credibility. *Digital Journalism*, *8*(4), 548–562. https://doi.org/10.1080/21670811.2020.1762102

Trafimow, D. (2013). The importance of surprising findings in psychology. *Frontiers in Psychology*, *4*(173), 1–2. https://doi.org/10.3389/fpsyg.2013.00173

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024

Tversky, A., & Kahenman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Turing, A. M. (1950). I—Computing machinery and intelligence. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

Waddell, T. F. (2018). A robot wrote this? *Digital Journalism*, *6*(2), 236–255. https://doi.org/10.1080/21670811.2017.1384319

Wang, S. (2021). Moderating uncivil user comments by humans or machines? The effects of moderation agent on perceptions of bias and credibility in news content. *Digital Journalism*, *9*(1), 64–83. https://doi.org/10.1080/21670811.2020.1851279

Weise, K., & Metz, C. (2023, May 9). When A.I. Chatbots hallucinate. *The New York Times*. https://www.nytimes.com/2023/05/01/business/ai-chatbots-hallucination.html

Xu, K., Chen, X., & Huang, L. (2022). Deep mind in social responses to technologies: A new approach to explaining the computers are social actors phenomena. *Computers in Human Behavior*, *134*, 1–13. https://doi.org/10.1016/j.chb.2022.107321

Downloaded from https://academic.oup.com/hcr/advance-article/doi/10.1093/hcr/hqad034/7329158 by guest on 28 January 2024