

Rise of Machine Agency: A Framework for Studying the Psychology of Human–AI Interaction (HAI)

S. Shyam Sundar

Media Effects Research Laboratory, Donald P. Bellisario College of Communications, Penn State University, University Park, PA 16802, USA

Advances in personalization algorithms and other applications of machine learning have vastly enhanced the ease and convenience of our media and communication experiences, but they have also raised significant concerns about privacy, transparency of technologies and human control over their operations. Going forth, reconciling such tensions between machine agency and human agency will be important in the era of artificial intelligence (AI), as machines get more agentic and media experiences become increasingly determined by algorithms. Theory and research should be geared toward a deeper understanding of the human experience of algorithms in general and the psychology of Human–AI interaction (HAI) in particular. This article proposes some directions by applying the dual-process framework of the Theory of Interactive Media Effects (TIME) for studying the symbolic and enabling effects of the affordances of AI-driven media on user perceptions and experiences.

Keywords: Source Interactivity, Machine Heuristic, Artificial Intelligence (AI), Algorithms, User Experience, Human–AI Interaction (HAI), Theory of Interactive Media Effects (TIME)

doi:10.1093/jcmc/zmz026

For much of its first two decades, this journal focused on “social science research on computer-mediated communication via the Internet, the World Wide Web, and wireless technologies.” In the Fall of 2013, the statement in the “About the journal” section was reworded to read “social science research on communicating with computer-based media technologies.” This change from “communication via” to “communicating with” technologies may seem subtle, but it signifies a profound shift in the study of computer-mediated communication (CMC).

In its early days, CMC research focused on documenting changes that occurred to interpersonal and group communications when they moved from face-to-face to mediated settings (Kiesler, Siegel, & McGuire, 1984). Scholars were particularly concerned that the quality of communication would suffer because non-verbal cues would be filtered out when interacting online (Walther & Parks, 2002). The

Corresponding author: S. Shyam Sundar; e-mail: sss12@psu.edu

Editorial Record: First manuscript received on 23 October 2018; Revisions received on 5 March 2019 and 1 July 2019; Accepted by Dr. Mike Yao on 16 August 2019; Final manuscript received on 7 September 2019

emphasis was clearly on human–human communication, and the technology of mediation was seen as a hindrance, or an uninteresting channel at best.

As media became more interactive however, there was a gradual shift in focus. Communication scholars began studying our interactions with the technologies themselves. Several studies documented our tendency to treat computers as if they are autonomous social actors (Reeves & Nass, 1996), to feel transported into artificially created mediated spaces (Lombard & Ditton, 1997) and to even become one with the interface of the technology as in the case of a cyborg (Biocca, 1997), among numerous other effects of interacting directly with computer-based media. This area of research tends to be categorized as human–computer interaction (HCI), and is sometimes contrasted with CMC in terms of the locus of users’ source orientation—while we orient to other human sources in CMC, we orient to the computer as the source in HCI (Sundar & Nass, 2000).

Such distinctions have blurred somewhat in the age of mobile and social media, as users seamlessly interact with both the interfaces and other humans, often leveraging interface features to augment direct individual interactions with media themselves as well as interpersonal, group and mass communications. Moreover, the technologies underlying these media have become active themselves, observing and proactively contributing to such interactions and communications. As a result, scholars have shifted their focus from the locus of communication to the “affordances”¹ of mediation technologies, by asking questions like: What can technology *afford*? How can technology enable human action, help humans, enhance humans, and how can we use technology for human needs and ends?

Perhaps the best example of an affordance that blurs the line between CMC and HCI is “source interactivity,” or the ability of users to serve as sources of communication (Sundar, 2007). At its core, source interactivity provides unprecedented agency to ordinary users by enabling them to not only customize information for themselves but also curate and create content for others—a privilege that until recently was the exclusive domain of journalists and other elites with access to media vehicles. As Bandura (2001) notes, personal agency gains in magnitude when individuals influence others in desirable and self-fulfilling ways (proxy agency). Feedback for one’s actions in social media provide constant reminders of this influence by way of metrics signaling the exercise of source interactivity by one’s contacts in the form of likes, comments and retweets. Such external validation is associated with psychological empowerment (Stavrositi & Sundar, 2012).

Aside from psychological benefits, auto-generated metrics can also have economic benefits. Number of page views, impressions and clicks are important determinants of payouts from self-service ad technology embedded in major search and social-networking sites (Subramanian, 2017). This has given rise to a new breed of social media celebrities, especially on YouTube, but also incentivized the creation of clickbait and sensational stories that are not always based on facts. The latter is associated with the spread of fake news on the Internet, with major efforts currently underway to develop automated solutions to tackle this problem (Pérez-Rosas, Kleinberg, Lefevre, & Mihalcea, 2017). We seem to be heading toward a future where autonomous bots are tasked with limiting the self-agency of bad human actors.

Rise of machine agency

While we may associate source interactivity and the consequent realization of self-agency with both the positive potential for empowerment and the negative potential for misinformation, we must recognize the growing power of the underlying technology. The technology that enables users to customize and thereby provide proxy agency to users is becoming increasingly capable of exerting its own agency, thanks to advancements in machine learning and artificial intelligence (AI). Machine learning is the ability of computing technology to identify patterns from data and infer underlying rules (e.g., Kaplan &

(Haenlin, 2019), as in deriving a set of linguistic, structural, source and network features that would help predict whether a given news story is real or fake. AI is the autonomous application of these rules by a system for adaptively achieving specific goals, such as making a decision or offering a recommendation, e.g., a plug-in that proactively alerts users about an incoming story in their social-media news feed as being fake. While the deployment of machine learning and AI to address the fake news problem is emergent, these technologies have been part of the media landscape for several years in the form of personalization systems that tailor media content based on a user's prior online actions. Examples include search engines that provide individualized search results, web portals that display only items of interest to the user, media services like Netflix that recommend certain movies over others, and advertisements that follow a user from site to site—all these are based on digital traces left by the user.

Such proactive inference of rules governing an individual's media behaviors by machines can understandably raise privacy concerns and threaten human agency. Personalization (whereby media systems covertly tailor content for users) would be less favored than customization (wherein the user performs the tailoring themselves). Sundar and Marathe (2010) tested this proposition with Google News (a news aggregation site) and found that it was true only for power users, i.e., those who are tech savvy and have higher levels of motivation, expertise, experience and efficacy in using information and communication technologies. Non-power users, on the other hand, preferred personalization over customization. Power users tend to assume low privacy as the default and therefore prefer to engage in self-tailoring rather than letting the system tailor media for them. Data showed that as power usage levels increased, perceived control increased in the customization condition, but decreased in the personalization condition.

Findings like this signal the essential tension between machine agency and human agency. While users appreciate—indeed welcome—the convenience of machines serving them, they are hesitant to cede decision-making control to them. As Rammert (2008) notes, machines differ in the degree to which they usurp agency, ranging from passive (which are completely driven from outside, e.g., hammer) and semi-active machines (which have some self-acting aspects, e.g., record player) to re-active (systems with feedback loops, such as a thermostat-driven climate control), pro-active (self-activating programs, e.g., car stabilization) and co-operative ones (distributed and self-coordinating systems such as smart homes). Technologies in the middle of this continuum strike a balance by providing proxy agency to users, performing actions at the user's command, including apps that help individuals manage their emotions (e.g., face punching app) as well as mobilize others (e.g., blog sites). Machines at the latter end of the continuum have grown by leaps and bounds in recent times with advancements in networking technologies and AI. We now have a number of technologies that proactively direct our attention with notifications and dictate our behaviors with suggestions, e.g., smartwatches that urge us to get up and walk and autonomous robots, including telepresence robots, that exude “apparent agency,” i.e., they are seen as acting on their own accord (Takayama, 2015). Ubiquitous computing and ambient intelligence technologies go a step further by rearranging our environments even without involving us in the process, e.g., smart homes.

Smart devices, smart homes, smart cars and smart cities are all premised on the notion of interoperability of different systems, communicating with each other on behalf of the user but with minimal direct involvement of the user. But, there is a downside: Remember the famous satirical video produced by ACLU in 2004 (<https://www.aclu.org/ordering-pizza>), entitled “Ordering pizza in 2015,” forecasting a future where a pizza delivery service would be plugged into your health, police, shopping and financial records, allowing it to calibrate the cost of your pizza based on the added risk of consuming high-fat foods to your health (as per your medical history) and also the added risk posed to the driver who delivers the pizza (premised on the crime profile of your neighborhood)? This is now a reality in ways

more hidden than was imagined 15 years ago. Surveillance of humans by ubiquitous computing systems is so common now that we have come to expect recordings and digital traces for solving virtually all crimes and misdemeanors (e.g., Clark, 2013). We have also come to the realization that privacy is no longer the default assumption; in fact, one has to opt out of being tracked in order to preserve one's privacy (e.g., Schuppe, 2018). This has resulted in a cottage industry of technologies that afford privacy, such as virtual private networks (VPN), incognito mode and anonymous messaging apps. Users value apps that promise data minimization, as evidenced by the phenomenal success of Snapchat, whose principal feature is that photos and messages disappear and become inaccessible after a few seconds. This affordance of ephemerality is particularly welcome in a technology environment where our digital traces are preserved into eternity. Young users have been quick to adopt Snapchat in large numbers because they do not have to worry about the accountability that stems from their posts being recorded and visible to unknown others (Xu, Chang, Welker, Bazarova, & Cosley, 2016). More broadly, it signals a human desire for technologies that limit machine agency and a shift toward users asserting their personal agency, with companies like Mozilla making individual user privacy a core aspect of their mission.

It may seem ironic that we are turning to machines for limiting machine agency and reclaiming human agency, but it signals an emergent collaboration between humans and machines in negotiating the type and degree of agency. This collaboration rests on a nuanced understanding of the various ways by which machine agency can enhance human agency and the ways by which it may threaten it.

How does machine agency enhance human agency? For starters, it can collect more data and perform more complex analytical operations than is humanly possible, allowing individuals to make better, data-driven decisions. Machines offer unprecedented levels of service and convenience in the form of recommendation agents by providing just-in-time information, as in a car navigation system, thus freeing up cognitive and physical resources. In the media context, they afford source interactivity, as discussed earlier, so that users can customize, curate and create media content. More generally, they afford greater user control over interactions, including subversive technologies like anonymous messaging apps that allow users to subvert surveillance and avoid leaving digital traces. Machines in the form of smart speakers (e.g., Amazon Echo) can aid humans by improving their communication skills, facilitating social interactions and reducing their communication apprehension. They can serve as reliable companions, just like robots—always dependable, never tiring of answering the same question over and over again—helping humans by providing companionship and cope with disorders such as dementia. Furthermore, they can empower humans through shared cognition and augmenting human abilities in a number of domains, from chess playing to travel booking (e.g., O'Reilly, 2017).

Even though this potential for enhancing human agency is the driving force behind the industry push toward AI, the popular narrative is quite the opposite. The media discourse is dominated by fears of automation and consequent job losses. With increasing interoperability among systems, machines are more in control than ever before, often leaving the human out of the decision-making loop. When mobile apps communicate with our laptop via the cloud and automatically update our interaction history to provide highly personalized services, we tend to be concerned that the boundaries between systems are fuzzy and porous. Humans fear that machines are getting so complex that it is often difficult to exercise human oversight even when interfaces sometimes offer them. Such concerns are not limited to nuclear plants, automated factories, self-driving cars and robots, but apply also to everyday media and communication technologies. The privacy settings offered by Google and Facebook, for example, can pose quite some challenges to users as they have to think through a variety of scenarios in which they weigh the benefits of exposing their queries and posts to their networks against the costs to their privacy. Users often realize the pitfalls of their settings when their privacy or security is compromised,

as when a social media post is seen by those who were not meant to see it. Under such circumstances, users feel hapless in the face of the various machinations of the algorithms that implement rules without taking into account the nuances of human nature and relationships. A similar scenario gets played out when bots programmed to be bad actors enter the online communication environment, as happened with Twitter bots targeting users with many followers through replies and mentions and thereby persuading their human followers to unwittingly reshare fake news in large numbers during the 2016 U.S. presidential elections (Shao, Ciampaglia, Varol, Yang, Flammini, & Menczer, 2018). More generally, as smart machines dominate our landscape, users are made to feel anything but smart. Several correlational studies have attempted to link the use of contemporary communication technologies with lower intelligence (Carr, 2010), lower self-efficacy, lower interpersonal skills and poorer mental health (Twenge, 2017). As the media-equation literature notes, users cannot help themselves but be social toward machines (Reeves & Nass, 1996), raising moral concerns about human attachment to non-human agents and inspiring a sub-genre of Hollywood movies such as *Her* and *Ex Machina*. In sum, increasing machine agency is seen as a threat to human agency in this counter-narrative.

Going forth, instead of simply documenting gain or loss of human agency, it would be fruitful for research and design to understand the trade-offs, and focus on strategies for negotiating agency between the human user and the intelligent machine. One strategy might be to seek user assent before taking decisions on behalf of the user. For example, in systems that personalize media for users, “reactive personalization” (obtaining user permission before providing personalized recommendations) would be preferable over “proactive personalization” (i.e., automatically pushing recommendations to users) because it is solicited by the user (Bellavista, Küpper, & Helal, 2008). Another strategy is to increase transparency pertaining to machine operations, e.g., how user data would be collected, stored and used (Chen & Sundar, 2018). Greater visibility of user input, rather than ambient collection of user data, is also associated with a better balance between human agency and machine agency (Jia, Wu, Jung, Shapiro, & Sundar, 2012). Machines may also be co-opted to provide media literacy to users by alerting them about the pitfalls of relying too much on automated recommendations and directives (Zhang, Wu, Kang, Go, & Sundar, 2014). That said, when the bandwidth of human information processing is clearly inadequate for sifting through all the data and information, algorithms could be deployed to help users make better inferences. For example, automated fact checkers (e.g., Hassan, Arslan, Li, & Tremayne, 2017) can detect and flag fake news in our online feeds more efficiently than humans, but without threatening human agency because they would be used only as a first line of defense and often in concert with human monitoring.

Machine agency in mediated communication

As newer media and communication technologies incorporate greater levels of machine learning, AI has become an integral part of the concept of mediated communication. Media scholars who are used to investigating the effects of messages to the exclusion of medium factors are having to account for the fact that increasingly an intelligent entity is mediating content, and that media are no longer mere uninteresting channels between human senders and receivers. For CMC scholars, mediation by AI adds an exciting set of new affordances that will likely redefine the future of mediated human–human communication. When Gmail autocompletes our messages or when Google Translate acts as the interpreter facilitating our conversation with a person using a different language, messages are proactively modified by AI technology. The deployment of such AI-mediated communication (AIMC) in CMC applications can profoundly affect interpersonal interactions and human relationships by raising questions about the human sender’s intentionality, authenticity and credibility (Jakesch, French, Ma, Hancock, & Naaman, 2019).

For HCI scholars, AI is the source of their interaction, much like programmed computers were in the 1990s, but with considerably more agency and decision-making abilities. We are moving toward an era where the socialness of our interaction with a computer is no longer an oddity but indeed a reasonable response because it possesses intelligence that is not just human-like, but surpasses human abilities. The explicit incorporation of social elements into the interfaces of contemporary media tools, such as virtual assistants and smart speakers, renders them even more worthy of human attributions and social responses.

Very soon, most HCI will be Human–AI interaction (HAI) whereby users will orient to the media source as if it is an intelligent entity that is capable of modifying content in unprecedented ways. In one respect, this is an extreme realization of the potential of interactivity. The technology of the medium can no longer be treated as a constant (as it currently is in most communication research), but as a variable, indeed a bundle of variables. With AI mediation, there is no such thing as medium-agnostic content, as content is profoundly shaped by the characteristics of the AI driving a given medium or media interface. Several scholars, especially McLuhan (1964), have long argued that medium changes the nature of content, but an AI-driven medium will make it too apparent to ignore. When we attempt to make content constant across media, the affordances of a given medium will invariably alter the content. For example, the same piece of fake news will appear different in different media because the algorithms embedded in each will shape the content based on its unique, distinct rules. Likewise, there will be no such thing as content-agnostic medium any longer because the manner in which AI-driven affordances manifest themselves is shaped by the content that is typically experienced within the medium. For example, we associate Alexa, the smart speaker, with a certain repertoire of content rendered in a certain way, quite different from other AI-driven media such as Google, the online search engine. Moreover, the application of an AI medium's algorithm to one genre of content (e.g., interactive news) is likely to alter content in a manner that is different from application of the same or similar algorithm to another genre (e.g., interactive movies). As a result, we cannot study technology independent of content. What we need going forth are interaction hypotheses crossing technological variables with content variables.

Theorizing human interaction with agentic machines

Just as media content is “infinitely describable” (Cappella & Street, 1989), media technology is similarly multivariate. HCI and CMC theories focusing on specific variables, ranging from anonymity and customization to depersonalization and interactivity, could be employed to study emergent media, but with a focus on the specific affordances of AI and concerns surrounding the rise of machine agency. The Theory of Interactive Media Effects ([TIME]; Sundar, Jia, Waddell, & Huang, 2015) is ideally suited for this purpose because it focuses on the effects of technological affordances in digital media, which are the primary independent variables of interest. Specific affordances of AI-based media could be studied for both their perceptual effects and experiential effects via two distinct sets of psychological mechanisms proposed by the theory. The symbolic aspects of AI can be explored via the cue route and the enabling aspects of AI via the action route of TIME, as described below.

Cue route of human–AI interaction

AI as symbol

Merely identifying AI as the locus of user interactions can serve as a cue for triggering a variety of heuristics based on stereotypes about the operation of machines. Layperson perceptions of machines is that they are rule-governed, precise, accurate, objective, neutral, infallible, and when entrusted with private information, do not gossip like some humans (Sundar & Kim, 2019). On the negative side of

the ledger, machines are thought to be mechanistic, unyielding, unemotional, cold, transactional and prone to being hacked. As Lee (2018) discovered, users tend to attribute decisions made by algorithms to their efficiency and objectivity, which render them fit for mechanical tasks but unfit for “human tasks” that involve subjective judgments and emotional capabilities. In this way, both positive and negative stereotypes of machines will be invoked when an algorithm or bot is the source of interaction in HAII. These stereotypes form the basis of “machine heuristic,” which is a mental shortcut whereby we attribute machine characteristics when making judgments about an interaction. According to TIME, 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.

Perceiving visible attributes of AI

Aside from identity of source as AI, its visible attributes, such as system transparency (as in explainable AI, or XAI), can trigger positive heuristics and thereby lead to better user engagement. Eslami et al. (2015) discovered that a majority of Facebook users were unaware that an algorithm curated their news feeds, so they developed a system called FeedVis that showed study participants the difference between a curated and a non-curated news feed. Most respondents were initially surprised and upset because their expectations appeared to be violated by the deletion from their news feed of posts by certain friends. But, after they had an opportunity to compare the two feeds and felt more knowledgeable about the algorithm, they were generally satisfied, even appreciative of its filtering function. Greater algorithm awareness also resulted in more informed use and greater satisfaction with the Facebook news feed. This mirrors findings in the literature on recommendation systems and Internet of things where transparency of systems have contributed to greater trust and positive user experience (Chen & Sundar, 2018; Zhang & Sundar, 2019).

More generally, disclosure of AI identity and characteristics of its operation can shape the quality of HAII by triggering corresponding cognitive heuristics about the underlying nature of AI. The machine heuristic, described earlier, may result in positive or negative expectations and experiences depending upon the appropriateness of applying machine attributes to the activity at hand. For instance, as a user of a social networking site, you might trust and appreciate the nudge you receive from it to wish a friend well on their birthday, but as a recipient of a birthday wish from a friend in the network, you would likely resent it if it seems clearly automated. For some AI-enabled tasks, however, mechanistic operations are appropriate and will lead to greater trust, as in safeguarding one’s personal financial information (Sundar & Kim, 2019).

Prior experience as guide

The psychological effects of reliance on a heuristic are also dependent upon the user’s prior experience with AI and their general digital literacy. For example, when the interface of a fake news detection system emphasizes the algorithmic nature of its operations, it could cue the machine heuristic and thereby promote “automation bias” (Mosier, Skitka, Burdick, & Heers, 1996), which is the tendency to overtrust machines and underestimate human acumen to perform the same tasks. As a result, users are less likely to be vigilant about fake news, preferring instead to cognitively outsource the task to the trusted machine. However, if a user finds out the hard way that the algorithm failed to flag fake news, even if only occasionally, cueing the machine heuristic may result in negative reactions, such as algorithm aversion (Dietvorst, Simmons, & Massey, 2015), the tendency to prefer human judgments over algorithmic decisions even when it is suboptimal. Such reactions arising from heuristics based on a user’s prior experience will shape perceptions of the AI-driven medium and thereby govern user experience (see top pathway in Figure 1).

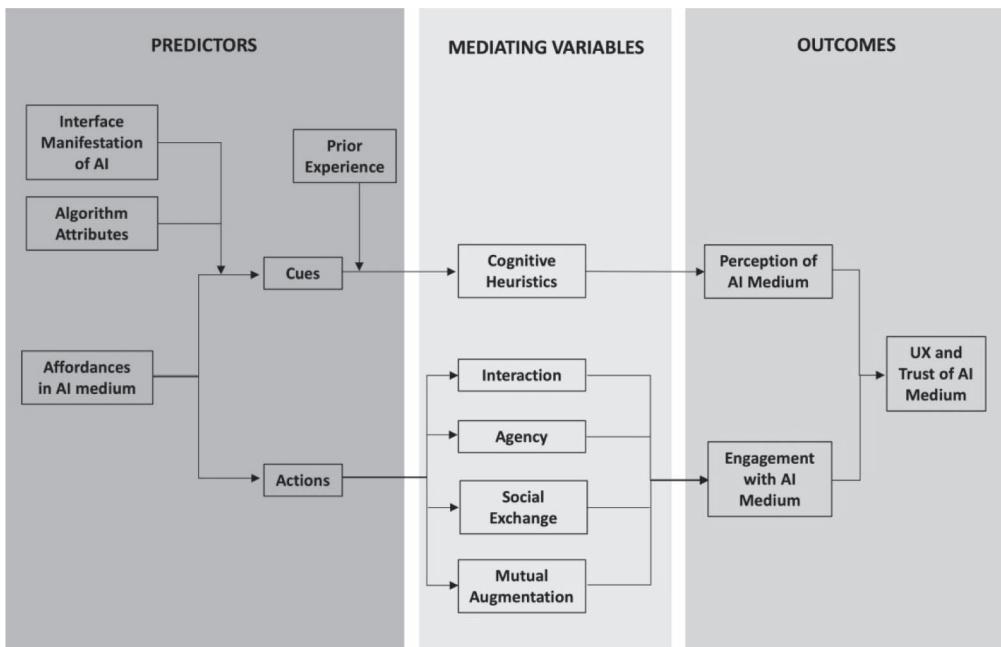


Figure 1 HAIITIME model: An adaptation of the Theory of Interactive Media Effects (TIME) for the Study of Human–AI Interaction (HAI).

Perceiving underlying attributes of AI

Aside from individual differences in prior experience with algorithms and general algorithm awareness based on visible attributes, specific underlying attributes of algorithm functions or operations can also serve as cues for triggering cognitive heuristics that shape expectations, perceptions and experiences. Users are known to have “folk theories” about how social media algorithms function “under the hood” (French & Hancock, 2017). Interview studies have shown that social media users hold several mental constructions about how algorithms work and the rules underlying their curation function—what Bucher (2016) refers to as “algorithmic imaginary”—based on their subjective appraisal of certain specific experiences with the medium. These “theories” can be the bases for cognitive heuristics that are triggered by particular features or outcomes of an AI affordance. For example, a common theory is that the more popular a post (as measured by number of likes, comments, retweets) and/or the person posting it (number of followers), the more prominent its appearance on your feed (Eslami et al., 2016). Belief in this theory is known to dictate user reactions to the algorithm, including feelings of resignation (DeVito, Gergle, & Birnholtz, 2017), even though, in reality, the algorithm may not be applying this popularity-based principle for its filtering function. A more transparent system would clearly be better in this case, for fostering more accurate perceptions of the algorithm’s functioning as well as for justifying user reactions toward it. If indeed the curation is based on popularity, then indicating it on the interface, by way of interface cues (such as number of likes or star ratings) can serve to trigger the appropriate heuristic—in this case, the “bandwagon heuristic” from TIME—and thereby shape psychological effects that are premised on truth rather than conjecture.

Toward this end, Alvarado and Waern (2018) propose the concept of Algorithmic Experience (AX) as an analytic framework for making user interactions with algorithms more explicit. For an ideal AX, users ought to be aware of how the algorithm functions and what it tracks in order to provide personalized services. Furthermore, they should be able to manage, corroborate and regulate its profiling, with the option of directing its future behaviors (for avoiding or emphasizing certain kinds of outcomes). The two key features of AX—user awareness and user control—are the hallmarks of successful user experience with personalization services, as evidenced by several recent studies (e.g., Zhang & Sundar, 2019; Chen & Sundar, 2018). Providing explanations about the how, why and what an algorithm does can increase algorithmic transparency (Rader, Cotter, & Cho, 2018) whereas not doing so can lead to anxiety and trial-and error reverse-engineering (Jhaver, Karpfen, & Antin, 2018). When viewed with the lens of TIME, the former would cue heuristics (e.g., control heuristic) that can shape users' psychological experience of the medium whereas the latter would give rise to folk theories and algorithmic imaginaries.

In sum, the cue route of TIME predicts that affordances in AI medium can trigger cognitive heuristics not only by advertising their existence on the interface of the medium, but also by providing interface indicators of—and/or clues in their output about—their *modus operandi*. These heuristics in turn shape users' psychological responses to the AI medium. As Kim (2016) discovered in his study of Internet of Things (IoT) devices, providing source cues by having each device communicate in a unique voice and designating a device as a specialist (rather than generalist) can positively affect user experience by triggering heuristics pertaining to social presence and expertise respectively. In this way, the cue route is primarily concerned with transparency and visible aspects of the AI system powering the medium.² Specific research questions about how particular interface manifestations of AI and evident attributes of the algorithm provide distinct cues (sometimes inadvertently), and how these cues interact with prior experience in triggering positive and negative heuristics can advance our knowledge about the mechanisms underlying the perceptual effects on user trust and experience, with clear implications for design of interfaces for AI-based media.

Action route of human–AI interaction

Human–AI collaboration effects

When the user engages the affordances of the AI system and provides their input, as in exerting control over an algorithm (the second aspect of AX), the nature of HAI will be shaped by the quality and outcomes of the collaboration that ensues (bottom pathway of Figure 1). Social robots, online chatbots, smartphone voice assistants and several other AI-driven media products are conversation agents that depend on user responses for interaction. They serve as social partners to human users, with anthropomorphic qualities (such as the human voice of smart speakers) potentially creating feelings of homophily among users and leading to greater socialness in the exchange. Given the interactive nature of these systems, users can either choose to be guided by the machine and conform to its directives or exert control over it by customizing settings. In this way, the action route is premised on the availability of volitional control for users.

Aside from negotiating agency, the action route of HAI holds out potential for human–machine synergy—just like in the interpersonal context, the more relational a smart technology, the higher the mutuality and interdependence (Rusbult & Arriaga, 1997), leading to a tightly coupled collaboration. For example, personalization systems or smart devices that are overt about their collection of user information or provide users an opportunity to give their assent (Bellavista et al., 2008) would be considered more collaborative than those that covertly collect user data and proactively provide

personalized services without user consent or authorization. In sum, the provision of interaction and the promise of greater user agency are key enabling aspects of the action route that will foster greater user engagement with AI media.

Costs and benefits of human–AI collaboration

All this does not mean that more user action is necessarily better. After all, the purpose of deploying tools of AI is to outsource human tasks and decisions to machines so that they can enhance the comfort and convenience of humans. As social exchange theory (Roloff, 1981) would predict, their success depends on their ability to advance the interest of the user at minimal cost. In this context, user actions on the interface, such as searching, choosing settings and making decisions, are costs against which the benefits of AI media, such as tailored content and convenience, would be assessed.

Aside from social exchange, the action route proposes the possibility of mutual augmentation of human users and AI systems. From optimizing routing for Uber drivers to suggesting listing price for Airbnb hosts, algorithms augment human decision-making by extending the scope, range and speed of information processing. Facebook notifications that remind and nudge us to wish our friends on their birthdays and online dating apps that provide us with data-driven matches are powered by algorithms that help augment our social lives. In organizations, algorithms help screen job candidates and suggest strategies for optimizing resources. In all these examples, AI is a decision aid, not the decision-maker, yet it plays a transformational role in augmenting human ability for decision-making. Scenarios like this are more common than the doomsday scenarios of naysayers who fear widespread job losses for humans (Vincent, 2018). As Jarrahi (2018) notes, the more pragmatic perspective highlights the complementarity of AI and humans. “AI can extend humans’ cognition when addressing complexity, whereas humans can still offer a more holistic, intuitive approach in dealing with uncertainty and equivocality” (p. 577). This notion of “intelligence augmentation” or “intelligence amplification” (IA) has been around since the beginning of AI (Rheingold, 1985), but modern-day interactive interfaces have realized this vision by providing affordances that facilitate seamless collaborations between humans and intelligent machines.

Human–AI synergy

While much has been said and written about AI augmenting humans, relatively little attention is paid to the possibility of—indeed, the very real necessity for—augmenting AI systems with human abilities and concerns. In order for a healthy symbiotic relationship between humans and AI, users ought to be provided more opportunities for directing the manner in which algorithms function and cater to their specific needs. When Alexa makes meta-cognitive assertions like “Sorry, I don’t know that,” users ought to be provided a mechanism by which they can enable Alexa to acquire this kind of knowledge. This applies not only to expanding the scope of services by AI to humans but also to improving the accuracy of AI-driven insights. Otherwise, users will be inadequately served at a minimum and outright unhappy at the extreme, giving rise to phenomena such as “algorithmic anxiety” discovered by Jhaver et al. (2018), who found that Airbnb hosts felt hapless not knowing how the algorithm evaluated them and why they were ranked at a certain spot on search outputs, or given certain pricing suggestions. Providing complete transparency to the hosts may not be a viable solution in this case, as it could result in gaming of the system and ultimately reduce the credibility of Airbnb, but users should be provided an avenue for “ground-truthing” the algorithm with their individual experiences. Providing such an opportunity to users is akin to feeding accurate and comprehensive training data to machine-learning algorithms in order to ensure higher accuracy, but applied to the case of individual users, so that each person’s unique circumstances can be factored into the inferences made by the AI system. Such opportunities

for individual users to play a part in training algorithms can realize the full potential of the enabling aspect of AI-driven media, while also compensating for the lack of transparency, which, as discussed earlier, is critical for psychological effects via the cue route. Of course, not all users will be interested or efficacious enough to engage in training algorithms. Some may even resent the effort needed to do so. But, human input into AI systems need not necessarily come from users. Human agency can be incorporated into AI media by explicitly incorporating rules built by humans (as in supervised machine learning) and/or by having other humans serve as co-authors. At least one study (Waddell, 2019) has found that news attributed to machine and human sources in tandem is rated higher in credibility than the same news attributed to either source in isolation.

In sum, the action route of TIME predicts that the various user actions afforded by AI interfaces can dictate user engagement and experience based on the extent to which they allow users to interact with the system, assure them of human agency, provide tangible benefits and offer avenues for mutual augmentation. This route is dictated by the nature of the collaboration between human users and AI systems, quite unlike the cue route, which is based on human perception of the manifestation of AI systems. It can also serve to mitigate perceptual concerns arising from lack of algorithmic transparency, which is not always possible (either because of concerns of hacking or because the AI is a black box based on self-learning underlying patterns that are so complex as to be a mystery even to the designer). As such, the action route is likely to be more effortful than the cue route and determine user outcomes based on the level of user engagement or involvement. If dual-process models in social psychology are any guide, trust in AI systems built via the action route is likely to be more robust than that via the cue route, based as it is on deeper involvement with the algorithms and perceived understanding of their functions, and therefore more resilient to change even under circumstances of failure.

Conclusion

The future of media lies in synergistic systems that deftly leverage and combine the strengths of both machine agency and human agency. Whether it is mutual augmentation, trade-offs in human vs. machine agency, social exchange between the user and the medium or simply sustained interaction, the provision of action to the user will enhance user engagement with the medium and thereby shape user experience and trust in AI-based media. Identifying important AI affordances, salient cues and collaborative actions, as well as the mechanisms by which they affect user experience and trust in AI, both separately and interactively, will be important for future research. Empirical explorations of the effects of these cues and actions will shed light on the co-creation of reality by AI-embedded media and their users, thereby enhancing our understanding of the social and psychological consequences of emerging communication technologies in the future.

Endnotes

1 By “affordance,” I mean “action possibility” suggested by environmental stimuli (Gibson, 1977), which, when applied to human interactions with technology, refers to properties of the system that suggest ways in which it could be operated (Norman, 1988). Any given feature of a system can afford different actions depending upon how the users engage with it (Treem & Leonardi, 2012) and any given affordance can be realized with more than one feature. This does not mean that an affordance is purely perceptual, but rather a relation between the material features of the medium and user actions that are governed by them (Evans, Pearce, Vitak, & Treem, 2017). In practical terms, when applied to the concept of interactivity, this would mean assessing the degree to which the interactive potential

- of an interface was realized by users. Therefore, any reasonable investigation of an affordance would have to name some features and identify one or more proximate consequences of those features that describe user actions. These consequences in turn could be examined for their effects on psychological outcome variables of interest. For example, the affordance of “source interactivity” identifies some features (e.g., commenting function, blogging tool) as leading to the proximate consequence of acting as a communication source or expressing oneself (“self-expression”) before it is tied to outcome variables such as engagement. Therefore, the two variables in the first part of this path model, provision of blog tool (as an interface feature) and realization of self-expression (by the user), together constitute the *affordance* of source interactivity whereas the second part or path, from self-expression to engagement, constitutes the *effect* of source interactivity.
- 2 It must be noted that the cues here refer to heuristic value of the interface manifestation of AI and algorithm attributes of the AI medium, as discussed in the preceding paragraphs (e.g., provision of algorithmic curation; bandwagon metrics), and not the signaling of the possibility of user action by the affordances of an AI medium. The latter notion of “cues” is true for all affordances, i.e., they all convey action possibilities to users. TIME (Sundar et al., 2015) posits that the mechanism by which users form impressions based on simply recognizing the existence of these affordances (cue route) is different from the mechanism by which they form attitudes by engaging with the affordances of the medium (action route). The box labeled “cues” in the cue route (top portion of Figure 1) refers to any and all salient attributes of the affordance, including of course its sheer existence. The “action” in the action route refers to any and all actions undertaken by the user when presented with the affordance.

Acknowledgments

The author thanks five anonymous reviewers, editor-in-chief Rich Ling and associate editor Mike Yao for their feedback and suggestions on earlier versions of this article.

References

- Alvarado, O., & Waern, A. (2018). Towards algorithmic experience: Initial efforts for social media contexts. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*. Paper No., 286. doi: [10.1145/3173574.3173860](https://doi.org/10.1145/3173574.3173860).
- Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, 52, 1–26. doi: [10.1146/annurev.psych.52.1.1](https://doi.org/10.1146/annurev.psych.52.1.1).
- Bellavista, P., Küpper, A., & Helal, S. (2008). Location-based services: Back to the future. *IEEE Pervasive Computing*, 7(2), 85–89. doi: [10.1109/mprv.2008.34](https://doi.org/10.1109/mprv.2008.34).
- Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication*, 3(2). doi: [10.1111/j.1083-6101.1997.tb00070.x](https://doi.org/10.1111/j.1083-6101.1997.tb00070.x).
- Bucher, T. (2016). The algorithmic imaginary: Exploring the ordinary effects of Facebook algorithms. *Information, Communication & Society*, 20(1), 30–44. doi: [10.1080/1369118x.2016.1154086](https://doi.org/10.1080/1369118x.2016.1154086).
- Cappella, J. N., & Street, R. L. Jr. (1989). Message effects: Theory and research on mental models of messages. In J. Bradac (Ed.), *Message effects of communication science* (pp. 24–51). Newbury Park, CA: Sage.
- Carr, N. (2010). *The shallows: What the Internet is doing to our brains*. New York: W. W. Norton & Company.
- Chen, T.-W., & Sundar, S.-S. (2018). This app would like to use your current location to better serve you: Importance of user assent and system transparency in personalized mobile services. *Proceedings of*

- the 2018 CHI Conference on Human Factors in Computing Systems (CHI'18)*, Paper No. 537. doi: [10.1145/3173574.3174111](https://doi.org/10.1145/3173574.3174111).
- Clark, M. (2013, April 18). Security cameras were key to finding Boston bombers. *Stateline*. Retrieved from <https://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2013/04/18/security-cameras-were-key-to-finding-boston-bombers>.
- DeVito, M. A., Gergle, D., & Birnholtz, J. (2017). Algorithms ruin everything: # RIPTwitter, folk theories, and resistance to algorithmic change in social media. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 3163–3174. doi: [10.1145/3025453.3025659](https://doi.org/10.1145/3025453.3025659).
- Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General*, 144(1), 114–126. doi: [10.1037/xge0000033.supp](https://doi.org/10.1037/xge0000033.supp).
- Eslami, M., Karahalios, K., Sandvig, C., Vaccaro, K., Rickman, A., Hamilton, K., & Kirlik, A. (2016). First I “like” it, then I hide it: Folk theories of social feeds. *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2371–2382. doi: [10.1145/2858036.2858494](https://doi.org/10.1145/2858036.2858494).
- Eslami, M., Rickman, A., Vaccaro, K., Aleyasen, A., Vuong, A., Karahalios, K., Sandvig, C. (2015). I always assumed that I wasn’t really that close to [her]: Reasoning about invisible algorithms in news feeds. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems (CHI’15)*, 153–162. doi: [10.1145/2702123.2702556](https://doi.org/10.1145/2702123.2702556)
- Evans, S. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2017). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, 22, 35–52. doi: [10.1111/jcc4.12180](https://doi.org/10.1111/jcc4.12180).
- French, M., & Hancock, J. (2017). What’s the folk theory? Reasoning about cyber-social systems. SSRN. doi: [10.2139/ssrn.2910571](https://doi.org/10.2139/ssrn.2910571).
- Gibson, J. J. (1977). The theory of affordances. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, and knowing: Toward an ecological psychology* (pp. 67–82). Hillsdale, NJ: Lawrence Erlbaum.
- Hassan, N., Arslan, F., Li, C., & Tremayne, M. (2017). Toward automated fact-checking: Detecting check-worthy factual claims by ClaimBuster. *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’17)*, 1803–1812. doi: [10.1145/3097983.3098131](https://doi.org/10.1145/3097983.3098131)
- Jakesch, M., French, M., Ma, X., Hancock, J., & Naaman, M. (2019). AI-mediated communication: How the perception that profile text was written by AI affects trustworthiness. *Proceedings of the 2019 Conference on Human Factors in Computing Systems Proceedings (CHI’19)*. Paper No., 239. doi: [10.1145/3290605.3300469](https://doi.org/10.1145/3290605.3300469).
- Jarrahi, M. H. (2018). Artificial intelligence and the future of work: Human–AI symbiosis in organizational decision making. *Business Horizons*, 61(4), 577–586. doi: [10.1016/j.bushor.2018.03.007](https://doi.org/10.1016/j.bushor.2018.03.007).
- Jhaver, S., Karpfen, Y., & Antin, J. (2018). Algorithmic anxiety and coping strategies of Airbnb hosts. *Proceedings of the 2018 Conference on Human Factors in Computing Systems Proceedings (CHI’18)*, Paper No, 421. doi: [10.1145/3173574.3173995](https://doi.org/10.1145/3173574.3173995).
- Jia, H., Wu, M., Jung, E., Shapiro, A., & Sundar, S. S. (2012). Balancing human agency and object agency: An in-depth interview study of the Internet of things. *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (Ubicomp’12)*, 1185–1188. doi: [10.1145/2370216.2370470](https://doi.org/10.1145/2370216.2370470).
- Kaplan, A., & Haenlein, M. (2019). Siri, Siri, in my hand: Who’s the fairest in the land? On the interpretations, illustrations, and implications of artificial intelligence. *Business Horizons*, 62(1), 15–25. doi: [10.1016/j.bushor.2018.08.004](https://doi.org/10.1016/j.bushor.2018.08.004).
- Kiesler, S., Siegel, J., & McGuire, T. W. (1984). Social psychological aspects of computer-mediated communication. *American Psychologist*, 39(10), 1123–1134. doi: [10.1037/0003-066x.39.10.1123](https://doi.org/10.1037/0003-066x.39.10.1123).

- Kim, K. J. (2016). Interacting socially with the Internet of things (IoT): Effects of source attribution and specialization in human–IoT interaction. *Journal of Computer-Mediated Communication*, 21(6), 420–435. doi: [10.1111/jcc4.12177](https://doi.org/10.1111/jcc4.12177).
- Lee, M. K. (2018). Understanding perception of algorithmic decisions: Fairness, trust, and emotion in response to algorithmic management. *Big Data & Society*, 1–16. doi: [10.1177/2053951718756684](https://doi.org/10.1177/2053951718756684).
- Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication*, 3(2). doi: [10.1111/j.1083-6101.1997.tb00072.x](https://doi.org/10.1111/j.1083-6101.1997.tb00072.x).
- McLuhan, M. (1964). *Understanding media*. New York: Signet.
- Mosier, K. L., Skitka, L. J., Burdick, M. D., & Heers, S. T. (1996). Automation bias, accountability, and verification behaviors. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 40(4), 204–208. doi: [10.1177/154193129604000413](https://doi.org/10.1177/154193129604000413).
- Norman, D. (1988). *The psychology of everyday things*. New York: Basic Books.
- O'Reilly, T. (2017). *WTF? What's the future and why it's up to us*. New York: Harper Business.
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. *ArXiv*. Retrieved from <https://arxiv.org/abs/1708.07104>.
- Rader, E., Cotter, K., & Cho, J. (2018). Explanations as mechanisms for supporting algorithmic transparency. *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Paper No. 103. doi: [10.1145/3173574.3173677](https://doi.org/10.1145/3173574.3173677).
- Rammert, W. (2008). Where the action is: Distributed agency between humans, machines, and programs. In U. Seifert, J. H. Kim & A. Moore (Eds.), *Paradoxes of interactivity* (pp. 62–91). Bielefeld, Germany: Transcript.
- Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, television, and new media like real people and places*. New York: Cambridge University Press.
- Rheingold, H. (1985). *Tools for thought: The history and future of mind-expanding technology*. Cambridge, MA: MIT Press.
- Roloff, M. (1981). *Interpersonal communication: The social exchange approach*. Beverly Hills, CA: Sage Publications.
- Rusbult, C. E., & Arriaga, X. B. (1997). Interdependence theory. In S. Duck (Ed.), *Handbook of personal relationships: Theory, research and interventions* (pp. 221–250). Hoboken, NJ: John Wiley & Sons Inc.
- Schuppe, J. (2018, July 30). Facial recognition gives police a powerful new tracking tool. It's also raising alarms. *NBC News*. Retrieved from <https://www.nbcnews.com/news/us-news/facial-recognition-gives-police-powerful-new-tracking-tool-it-s-n894936>.
- Shao, C., Ciampaglia, G. L., Varol, O., Yang, K.-C., Flammini, A., & Menczer, F. (2018). The spread of low-credibility content by social bots. *Nature Communications*, Article No., 9, 4787. doi: [10.1038/s41467-018-06930-7](https://doi.org/10.1038/s41467-018-06930-7).
- Stavrositu, C., & Sundar, S. S. (2012). Does blogging empower women? Exploring the role of agency and community. *Journal of Computer-Mediated Communication*, 17, 369–386. doi: [10.1111/j.1083-6101.2012.01587.x](https://doi.org/10.1111/j.1083-6101.2012.01587.x).
- Subramanian, S. (2017, February 15). Inside the Macedonian fake-news complex.” *Wired*. Retrieved from <https://www.wired.com/2017/02/veles-macedonia-fake-news/>.
- Sundar, S. S. (2007). Social psychology of interactivity in human–website interaction. In A. N. Joinson, K. Y. A. McKenna, T. Postmes & U.-D. Reips (Eds.), *The Oxford handbook of Internet psychology* (pp. 89–104). Oxford, England: Oxford University Press.
- Sundar, S. S., Jia, H., Waddell, T. F., & Huang, Y. (2015). Toward a theory of interactive media effects (TIME): Four models for explaining how interface features affect user psychology. In S. S. Sundar

- (Ed.), *The handbook of the psychology of communication technology* (pp. 47–86). Malden, MA: Wiley Blackwell.
- Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. *Proceedings of the 2019 Conference on Human Factors in Computing Systems Proceedings (CHI' 19)*, Paper No, 538. doi: [10.1145/3290605.3300768](https://doi.org/10.1145/3290605.3300768).
- Sundar, S. S., & Marathe, S. S. (2010). Personalization vs. customization: The importance of agency, privacy and power usage. *Human Communication Research*, 36, 298–322. doi: [10.1111/j.1468-2958.2010.01377.x](https://doi.org/10.1111/j.1468-2958.2010.01377.x).
- Sundar, S. S., & Nass, C. (2000). Source orientation in human–computer interaction: Programmer, networker, or independent social actor? *Communication Research*, 27(6), 683–703. doi: [10.1177/009365000027006001](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). Malden, MA: Wiley Blackwell.
- Treem, J. W., & Leonardi, P. M. (2012). Social media use in organizations: Exploring the affordances of visibility, editability, persistence, and association. *Annals of the International Communication Association*, 36(1), 143–189. doi: [10.1080/23808985.2013.11679130](https://doi.org/10.1080/23808985.2013.11679130).
- Twenge, J. M. (2017, September). Have smartphones destroyed a generation? *The Atlantic*. Retrieved from <https://www.theatlantic.com/magazine/archive/2017/09/has-the-smartphone-destroyed-a-generation/534198/>.
- Vincent, J. (2018, April 3). AI and robots will destroy fewer jobs than previously feared, says new OECD report. *The Verge*. Retrieved from <https://www.theverge.com/2018/4/3/17192002/ai-job-loss-predictions-forecasts-automation-oecd-report>.
- Waddell, T. F. (2019). Can an algorithm reduce the perceived bias of news? Testing the effect of machine attribution on news readers' evaluations of bias, anthropomorphism, and credibility. *Journalism & Mass Communication Quarterly*, 96(1), 82–100. doi: [10.1177/1077699018815891](https://doi.org/10.1177/1077699018815891).
- Walther, J., & Parks, M. R. (2002). Cues filtered out, cues filtered in: Computer mediated communication and relationships. In G. R. Miller (Ed.), *The handbook of interpersonal communication* (pp. 529–563). Thousand Oaks, CA: Sage Publications.
- Xu, B., Chang, P., Welker, C. L., Bazarova, N. N., & Cosley, D. (2016). Automatic archiving versus default deletion: What Snapchat tells us about ephemerality in design. *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (CSCW'16)*, 1662–1675. doi: [10.1145/2818048.2819948](https://doi.org/10.1145/2818048.2819948).
- Zhang, B., & Sundar, S. S. (2019). Proactive vs. reactive personalization: Can customization of privacy enhance user experience? *International Journal of Human-Computer Studies*, 128, 86–99. doi: [10.1016/j.ijhcs.2019.03.002](https://doi.org/10.1016/j.ijhcs.2019.03.002).
- Zhang, B., Wu, M., Kang, H., Go, E., & Sundar, S. S. (2014). Effects of security warnings and instant gratification cues on attitudes toward mobile websites. *Proceedings of the 2014 Annual Conference on Human Factors in Computing Systems (CHI'14)*, 111–114. doi: [10.1145/2556288.2557347](https://doi.org/10.1145/2556288.2557347).