Volume 2, 2021 ISSN 2638-602X (print)/ISSN 2638-6038 (online)

# Human-Machine Communication

ISSN 2638-602X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

Human-Machine Communication (HMC) is an annual peer-reviewed, open access publication of the Communication and Social Robotics Labs (combotlabs.org), published with support from the Nicholson School of Communication and Media at the University of Central Florida. Human-Machine Communication (Print: ISSN 2638-602X) is published in the spring of each year (Online: ISSN 2638-6038). Institutional, organizational, and individual subscribers are invited to purchase the print edition using the following mailing address:

Human-Machine Communication (HMC)
Communication and Social Robotics Labs
Western Michigan University
1903 W. Michigan Ave.
300 Sprau Tower
Kalamazoo, MI 49008

Print Subscriptions: Regular US rates: Individuals: 1 year, $40.
Libraries and organizations may subscribe for 1 year, $75.
If subscribing outside of the United States, please contact the Editor-in-Chief for current rate. Checks should be made payable to the Communication and Social Robotics Labs.

# Human-Machine Communication

**Volume 2, 2021**

**Christoph Lutz**, BI Norwegian Business School (Norway)
**Rhonda McEwen**, University of Toronto (Canada)
**Yi Mou**, Shanghai Jiao Tong University (China)
**Peter Nagy**, Arizona State University (U.S.A.)
**Seungahn Nah**, University of Oregon (U.S.A.)
**Gina Neff**, University of Oxford (United Kingdom)
**Jochen Peter**, University of Amsterdam (Netherlands)
**Sharon Ringel**, Columbia University (U.S.A.)
**Astrid Rosenthal-von der Pütten**, RWTH Aachen University (Germany)
**Eleanor Sandry**, Curtin University (Australia)
**Mauro Sarrica**, Sapienza University of Rome (Italy)
**Megan Strait**, The University of Texas Rio Grande Valley (U.S.A.)
**Satomi Sugiyama**, Franklin University Switzerland (Switzerland)
**Sakari Taipale**, University of Jyväskylä (Finland)
**David Westerman**, North Dakota State University (U.S.A.)

# CONTENTS

HUMAN-MACHINE COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.1

# Moving Ahead With Human-Machine Communication

**Leopoldina Fortunati[1] and Autumn Edwards[2]**

1 Department of Mathematics, Computer Science and Physics, University of Udine, Udine, Italy
2 School of Communication, Western Michigan University, Kalamazoo, Michigan, United States of America

For this volume, we selected 10 articles we found groundbreaking for their innovative contribution to the field of human-machine communication (HMC). This field becomes increasingly interdisciplinary since it attracts contributions from disciplines such as sociology and anthropology, as well as communication and media; we would like to see even more philosophy and politics. Interdisciplinarity is important to building an integrated lens through which to look at this new field and to solidifying social sciences as an attractive intellectual home for studying HMC. The range of topics addressed in these 10 articles is wide, branching in many exciting directions. However, when considered holistically as a collection, the articles in this volume deal substantially with four layers of discourse, which are relevant to moving forward the HMC project. They are as follows: (1) updates to theoretical frameworks and paradigms (MASA; Lombard & Kun), (2) examination of ontologizing and prototyping processes (Etzrodt & Engesser; Banks et al.), (3) a critical turn regarding gender (Liu) and ability/disability (Davis; Denhert & Leach), and (4) exploration of HMC in organizational contexts (Gibbs et al.; Johanssen & Wang; Piercy & Gist-Mackey; Prahl & Van Swol).

## 1. Progressing on CASA With the MASA Paradigm and Discussing it From Another Angle

Innovation at the theoretical level is necessary to work out a new field of research such as HMC but it is understandable that at the beginning scholars try to use the theories that they have inherited from past or contiguous fields. There are a plethora of theories available today, starting from those developed to address the field of mediated communication, passing through telephone communication (fixed and mobile) and especially through computer-mediated communication. Among these theories, we observe that the most important, theoretical development in HMC has focused on CASA (Computers as Social

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

Actors), the influential paradigm proposed by Nass and colleagues in the 1990s (Nass et al., 1994; Nass et al., 1996; Nass et al., 1997). According to CASA, people treat computers in some of the ways they treat humans by mindlessly applying to them the same social rules they use in human-human interactions. In this volume of *Human-Machine Communication*, Lombard and Kun in their article, "Social Responses to Media Technologies in the 21st Century: The Media Are Social Actors Paradigm," revisit CASA by expanding it and launching a new paradigm: the Media Are Social Actors (MASA) paradigm. This paradigm is elaborated in a very sophisticated way by introducing the distinctions among social cues, the effects of social cues on medium-as-social-actor presence and social responses, and then by examining the role of individual differences and contextual factors. Lombard and Kun also propose consideration of mindless and mindful anthropomorphism as two complementary mechanisms for understanding MASA. Based on evolutionary psychology explanations for sociality, they offer nine formal propositions and suggestions for future research to test and apply MASA. In the previous volume of *Human-Machine Communication*, Gambino, Fox and Ratan (2020) first proposed an extension of CASA, arguing that humans do not necessarily implement social scripts associated with human-human interactions as predicted by CASA, but perhaps also social scripts that have been developed for interactions specific to media entities. Their extension enables us to explain previous dissonant findings and expands scholarship regarding HMC.

The problem first addressed by Reeves and Nass (1996) is the acknowledgment that the spontaneous and immediate response to media is faulty. Humans respond in social ways to mediated representations that mimic real life, but without exerting the cognitive effort to determine how to respond in the right way. There are two crucial questions that await empirical investigation. The first is: How long does this shift from perceiving media as real people, places, and things to representations of them take? The second is: What happens when people realize the true nature of the media? The problem of the delay of time in understanding the real nature of technology is an old one and is addressed both by CASA and also by the Uncanny Valley theory by Mori (1970).

It is curious that in front of all the theories elaborated to address mediated communication, the media equation and CASA framework (Nass et al., 1996; Reeves & Nass, 1996) is identified as the one that seems to be the most suitable for helping the scholarship deal with the world of HMC. Thus, it is relevant to understand why this framework is considered so useful for this new field of studies. Gambino et al. (2020) argued that CASA is well-suited to accompany scholars in the HMC field because it is based on humans' understanding of social interaction. With the purpose of enhancing usability, they continue, designers have tried to reduce the cognitive effort it takes to use digital technologies by activating users' existing mental models. Designers and engineers create interfaces that reflect patterns of human communication and imitate offline forms of social interaction (Nass & Brave, 2005; Shneiderman et al., 2017). In effect, CASA presents the great advantage of prefiguring the immediate, potential patterns of human attitudes and behaviors in the field of social robots and virtual assistants. We argue that, however, this will not be sufficient because the field of social robotics is not limited to designing technological artifacts equipped with some primary and secondary cues, but is complicated by the question of the anthropomorphism even at the highest degrees (see Lombard and Kun in this volume).

Reeves and Nass (1996) have explained that people treat media as humans because their "old brains" did not evolve in a way to immediately distinguish mediated representations from their real-life counterparts. They write,

> The human brain evolved in a world in which *only* humans exhibited rich social behaviors, and in a world in which *all* perceived objects were real physical objects. Anything that *seemed* to be a real person or place was real. During nearly all of the 200,000 years in which *Homo Sapiens* have existed, anything that acted socially really was a person, and anything that appeared to move toward us was in fact doing just that. Because there were absolute truths through virtually all of human evolution, the social and the physical world encouraged automatic responses that were, and still are, the present-day bases for negotiating life. . . . Modern media now engage old brains. (p. 12)

They frame their argument by adding two other important elements: the first is that this incorrect perception is not only a question of the old brain but also a question of users' inattention and, second, that when users regain possession of their reasoning abilities they understand that their perceptions were wrong. In their words, "people frequently live life mindlessly and with little introspection" (p. 12) and there is "good evidence that media are real first, and false only after we think about it" (p. 13). Gambino et al. (2020) and Lombard and Kun (in this volume) have already advanced substantially the field's relationship to and application of the CASA framework. Here we would like to add new paths of discourse by arguing that, although this explanation might seem correct in evolutionary terms, it should be reframed in the light of perspectives coming from the historical, sociological, and semiotic (and hermeneutic) disciplines that we will discuss here.

From a historical point of view, we recall that in the Western world people have experienced the diffusion of automata especially during the renaissance and modernity. These automata replicated human features, as in the case of the Writer, the Musician, and the Draftsman, the three automata-androids built by Jaquet-Droz in the second half of the 18th century. The automata presented the same problem as contemporary media: they implicated a *delay* only after which the public could understand the truth of the technological artifact or even come to terms with the impossibility on the part of the public to understand this. In other words, the public at first glance did not understand that the automata were artifacts and this was the outcome of a deception. The Catholic Church opened a controversy on their alleged, demonic aspect, which was attributed to the magic halo that surrounded them and that derived from this delay between apparent and real/revealed nature. Through this delay, according to the Catholic Church, passed the infringement conveyed by automata as their aims were to not only surprise but also to *trick* the public. The related debate was connected to the old issue of "authenticity." As discussed by Baron (2013) the issue of authenticity crosses the long story of robots especially in their representation as androids or gynoids.

If we look at this problem from a sociological perspective, it turns out that humans do not always pursue authenticity and grant themselves various degrees of freedom in playing with inauthenticity. However, the foundations of any society are built upon authenticity, a basis on which humans will to some extent derogate, albeit only tactically. The implicit social contract, which is at the basis of the functioning of society, organizations, institutions, and sociality and which functions for mutual consent, is also based on a high degree of authenticity of the individuals and their social identities. Thus, at a certain point, one has to come back to authenticity. As Seabright et al. (2021, p. 19) argued, humans are "willing to punish those whom they believe to be cheating or unjust to others, even at a cost to themselves and even if they themselves have not been directly harmed by the cheating or unjust act" (Fehr & Fischbacher, 2004; Fischbacher, 2001). In conclusion, our old brains already have experienced this problem of distinguishing simulation from reality, which is not a unique manifestation of contemporaneity but is much older.

Second, the development of our brains and cognitive abilities may not follow a linear growth path that facilitates more ready distinction between the seemingly real and the real. We have also to consider that newer changes to our apparatus of thought and perception could render us in some ways less rather than more skilled in dealing critically with personified objects. As Simmel argues, with modernity our old brains have weakened themselves by being increasingly exonerated to think. In his essay on "The Metropoles and the Life of Spirit" (1903/1995), he describes how in the modern world the technology embodied in objects of daily use exempts individuals from previously required gestures, skills, and efforts. While things embody an increasing number of cultural layers, people become less and less culturally equipped and capable and the dissonance that follows transforms the modalities and the very meaning of existence and experience. Let us consider, continues Simmel (p. 54), the amount of culture that has been incorporated in the last 100 years in the objects and the cultural progress of individuals in the same period of time—even only in the upper classes. From this comparison, a terrifying difference in growth emerges to the detriment of individuals and even a certain regress of their culture (in terms not of education, but sensitivity and ideality) to the extent that the individual is less and less capable of coping with the luxuriant development of objective culture. Simmel's vision has not remained alone. For example, Barthes (1957) argued that to understand society, it is necessary to focus on the myths and symbology attached to the objects of everyday life. Douglas and Isherwood (1979) pointed out that objects serve to build social alliances or alienation. For Appadurai (1986), objects have social lives, are socialized things, and have social potential. Latour (1994) stated that it is impossible to build the social only with the social, and sociologists must recognize the role of objects as mediators of social interaction. Finally, Karin Knorr Cetina (1997) observed in modern societies an "increased orientation toward objects as sources of the self, of relational intimacy, of shared subjectivity and social integration" (p. 32). In conclusion, objects, and particularly technologies, have been increasingly recognized as unintentional mediators of social relationships. These overall observations seem to be prophetic given that while we build increasingly "smart" and "intelligent" artifacts, human IQ seems to decrease in the most recent decades (Pietschnig & Voracek, 2015). To conclude, our old brains in the last century probably have experienced a process of downsizing, which, even if recent, must be taken into account in the general framework of our reasoning.

A third and maybe most important consideration coming from semiotics is that our old brains have experienced narration (Chatman, 1980) and play since humans began to communicate through language. Oral societies communicated using forms such as narratives, myths (in particular, cosmogonies), legends, fables, and so on (Ong, 1986). When societies shifted from orality to literacy, the structure and mechanisms of narration were incorporated in written texts, as well. Thus, reading is a new cognitive function in evolutionary terms, because it was invented only 5,500 years ago (Dehaene, 2007; Wolf, 2007). By contrast, although it is impossible to recompose the phylogenetic history of narrative behavior and to affirm its evolutionary origin in a totally founded way, the structure and mechanisms of narration are very old and readers' brains are used to them (Cometa, 2017; Tomasello, 2005). Aristotle considered fiction a serious matter because it is capable of extending knowledge, allowing us to imagine what can happen in the future, and to be ready and open to any event (Consoli, 2011). When readers read a text, they make a pact with the authors (Sartre, 1960, p. 46), which consists of willingly suspending their disbelief in what they are told and believing in the world that is narrated to them (Eco, 2016). As readers, our brains are accustomed to cooperating in textual actualization and to considering as true the fictional notions provided by the author (Eco, 1983). Those whom authors consider as their Model Readers are, therefore, required to make this considerable "pact." Users are required to experience simulation, identification, and empathy (e.g., Rizzolati and Sinigaglia, 2006) and to share or not in what the author proposes. They have to collaborate based on their real-world competence and, when they don't, must accept what is provided by the authors. Authors not only ask readers to pretend they know things about the real world that readers don't know, but also to believe that they should pretend to know things that don't exist in the real world (Iser, 1987).

In addition to fictionality, another secret of the attraction of the narrative text is that its format has a magnitude smaller than the boundless one of the world of reality. Narrative worlds have more defined boundaries and this allows us to be more comfortable than in the real world. Although between the real and narrative worlds there is a relationship of partial identity, conceivability, and accessibility, the fictional world is more conceivable and accessible than the real world. More important is that the possible world of fiction is not an entity detached from the reality that gave it birth, but is a real part of this same reality. Eco (1985) explained clearly that possible worlds as epistemic constructs are real insofar as they are embedded, beyond syntactically, in the real world that produces them. They are not parallel; they are one inside the other and each of them participates a little in the reality of its own container. Thus, fiction requires the revision of our relationship with the real world through the ability to move in and out of the text (Montani, 2014) or to apply the two narrative faculties: decoupling and simulation (Cometa, 2017). Considering the complex relationships between fiction and reality leads us to examine play. Narration has the same function as play: Children play with dolls, kites, or wooden horses to familiarize themselves with the physical laws and with the tasks and roles that one day they will have to perform seriously. Likewise, reading stories means playing a game through which you learn to make sense of the immensity of things that happen in the real world. This is the therapeutic function of fiction and the reason why people tell stories.

The shift that people have undergone in respect to media is that they have transferred to media their experiences regarding narration and, in general, the content that media convey, and began to read the media themselves as texts (Cosenza, 2009). The relationship, the implicit contract that there is between the author and the reader, has been transferred to the relationship between the user and the medium. This displacement is a mechanism that our brains use often to approach new things. They reduce the impact of novelty by creating an oblique relationship with the new thing. Thus, based on what we have said so far, the motivation for why we treat media as if they were humans may be different from that proposed by Reeves and Nass (1996), because other factors—such as automata, the noticeable slowing or decline of IQ, and especially fictionality and play—intervene in the framework, and some of them are old.

The fourth consideration comes from semiotics and lies with symbols. An alternative to CASA's proposition that humans mindlessly treat computers and robots as real people because their old brains have not evolved to immediately tell the difference is that computers and robots may be interpreted as symbolic humans, understood to *represent* but not to *be* human persons. In their figural "semiotic triangle," Ogden and Richards (1923) posit that meaning is worked out in the relations between *symbol* (an item used to represent other things or ideas), *reference* (the initial thought associated with the symbol), and *referent* (the actual thing or object of meaning), an idea comparable to C. Peirce's earlier (1839–1914) Sign Theory (Peirce, 1998). When people respond to a virtual assistant, for example, as if it were a human, they may interpret (reference) the agent (symbol) as directly "standing for" or "standing in for" a human person (referent).

Long before computers were social actors in the research paradigmatic sense, they were human social actors in the lifeworld. As detailed by Grier in *When Computers were Human* (2005), in the mid-1700s, astronomers attempted to use the laws of gravity to predict the return of Halley's Comet, but the calculations proved too onerous for a single scientist. This necessitated a division of mathematical labor resulting in the creation of a new occupational role: the computer, or person whose job was to perform scientific calculations by hand. At first, computers were predominantly young men. However, by the 19th century, women, people of color, polio survivors, Jews, and others who faced grim labor prospects were hired to reduce costs. There was a concomitant association of computers with women, in specific. According to Hicks (2018), pre-electronic computation jobs were quickly feminized because they were perceived as rote and de-skilled. Ultimately, the computers' roles and functions were replicated and off-loaded to electronic machines bearing their old job titles. A similar lineage can be traced in the case of robots. As it is well known, the word "robot," derived from the Czech word for slave (*robotnik*), was first used in 1921 by Karl Čapek, whose play *R.U.R.* was about machine men that were built to work factory assembly lines and later rebelled against their human masters. Even in its earliest usage, the robot symbolized in both form and function the human being whose labor was forced and whose servitude was involuntary. Thus, human beings were the first computers and, in a symbolic sense, human beings were also the first robots.

The symbolic treatment of some people as robots continues. Several recent news stories have focused on what it means to be treated as a machine in organizational spaces. For example, Saintato (2020) writing for *The Guardian* suggests that people are treated like robots when corporations apply the unprecedented efficiency of the digital and data revolutions to the human workforce in terms of time management, surveillance, behavioral profiling, performance quantification, and replacement with cheaper and newer models. We contend that if one wonders whether their job is at risk of replacement by a robot, they need mainly consider the degree to which they are presently treated as a robot. This is because robotization of the human workforce tends to directly precede automation processes, as we will explain below. The idea is widespread that automation could enter factories because workers themselves had become the Great Industry appendages of the machines. However, at cognitive and cultural levels the notion that a person can be replaced by a robot in industry and also domestic spaces demonstrates a certain symbolic equation of the robot and human and perhaps illuminates why people sometimes treat robots like humans. This symbolic transfer is evident in our classrooms when, for instance, women express dismay at their peers' sexist abuse of female-voiced virtual assistants: "Is this how you treat (or how you wish you could treat) the women in your life?" These students express the view that such AI systems obviously are meant to represent them, to symbolize them as human women, and that disrespect of the symbol damages also the object of its reference. For them, using polite language with Alexa or Siri does not arise from the mindless application of human scripts, but from the same process that encourages careful treatment of a flag to not denigrate the nation for which it stands. More generally, this perspective can be framed in light of what Danaher (2017), in the context of sex robots, terms the "symbolic-consequences argument." Although problematic aspects of robot symbolism are perhaps *removable* and *reformable* (Danaher, 2017), symbolism itself remains endemic to social relations with robots.

Like narration and play, symbols are ancient and are constitutive of what makes humans human in the first place. From the analog virtual realities of cave paintings and human sculpture to language, to the digitally-rendered mind spaces and anthropomorphized artifacts of current technology, the ability to create and use symbols, or to overlay abstract and fictional structures onto our physical reality, is a marker of human social evolution (Danaher, 2017, p. 8; Harari, 2016). When we interpret or respond to the symbol in light of what it represents, it is not necessarily because of a naïve conflation or a mindless attribution that they are the same. Rather, it is because we are used to this surrogating process that Sneddon (2015) has articulated as *taking symbols in virtue of what they represent* and which governs the relational aspects of human life (Sneddon, 2015 as discussed in Danaher, 2017, pp. 7–8). Therefore, whereas CASA's "old brains engage new media" formulation leads naturally to a focus on mindless versus mindful attribution processes, these hermeneutic and semiotic interpretations of robots/media as narrative texts and symbolic humans beg scholarly attention to issues of literacy and representation, respectively.

This discourse on why people treat media as humans is only sketched and needs to be further developed and integrated. Surely, two issues need future empirical research. The first is the need to understand what happens after people have realized their mistake (i.e., that machines are not humans). A partial response might arrive from semiotics: the cooperation that is required of users needs time to unfold. The text to be read proposes itself as a becoming, dynamic process: readers actualize over time successive portions and consistent macro propositions of it (Eco, 1985). Narration requires a progressive comprehension, which is the outcome of the reader's negotiation with the text (Cecchi, 2019). In the hermeneutics of the relationship between narration and the experience of time (Ricoeur, 1985/1988), the exemplary exercise of imagination accompanies the horizon of waiting (Jauss, 1967/1969). Again, users need time to make sense of any text.

The second open issue is to understand why among the various strategies that readers generally apply to a text we have often transferred to media the naïve one. Let us outline the problem. Readers apply to texts several strategies according to their skills. There is a "naive reading" when readers identify themselves completely with the story and with the characters and mistake the story they are reading for true (Eco, 1983). This is obviously a "perspective error": in this case, readers do not realize that the events and characters of the narrative work have a limited existence. Then, there is an "available reading" when readers are aware of the (partially or totally) fictitious nature of a narrative text, are willing to temporarily suspend their disbelief, to accept the story as if it were true, and also the rules of the game. They identify themselves with the story, experience the emotions that the text and the author want to arouse (the ability to identify with the story of the characters is necessary to the functions of evasion and catharsis of reading, which are among the main functions sought by literature). Finally, there is a "critical reading" when readers do not foresee the identification of the reader in the story they are reading; it is a second, specialized reading which aims to criticize and interpret a text (Eco, 1983). Probably users often apply to media a naïve reading because information and communication technologies are relatively new and users have a limited repertoire, understood as the ensemble of conventions necessary to establish a situation common to the artifact and to themselves, as well as limited mental schemes and scripts. It is not surprising if the first strategy adopted by users is to apply a naïve reading and the mental schemes and scripts that users have tested in human-human communication in the real world. But after a while, as Gambino et al. (2020) argued, things change.

There is another element of the CASA theory to correct. Reeves and Nass (1996) argue that "people's responses to media are fundamentally social and natural" (p. 251). In the next section, we will see how it is more appropriate to substitute the term natural with cultural.

## 2. Ontologizing and Prototyping Moves

The second strand addressed in this volume offers two relevant articles, which advance the debate, the first on ontology and the second on prototyping. Etzrodt and Engesser, in the article "Voice-based agents as personified things: Assimilation and accommodation as equilibrium of doubt," contribute to the debate on the ontology of humans and machines (as well as animals and simple objects) by proposing a sophisticated investigation of the nature of doubt regarding voice-based agents, based on Piaget's ontological classification of object-subject and thing-person, their associated equilibration processes, and the key factors of the situation: the user and the agent. They documented that the majority of their participants classified voice-based agents into personified things, although this classification remains fragile. The notion of personified things summarizes well how the boundaries between humans and robotic technologies are becoming less clear. Much of the ontological debate on the human-machine communication field has focused so far on the increasing blurring of the boundaries between the various elements of the relation: human and machine, for example (Edwards et al., 2019; Guzman, 2020). Recently, the epistemologist Nikhil Bhattacharya, in a path-breaking seminar, argued that if we want to really progress with this question, we need to focus on the two elements of the interaction: the human and the machine (robot). In front of their blurring boundaries, there is the need to trace a dividing line that remains essential between humans and robots. Sandry (2015) articulates the value of *blurred yet meaningful boundaries* by drawing on N. Katherine Hayles's (2005) posthumanist argument that:

> the 'boundaries are *both* permeable and meaningful,' such that humans can still be regarded as 'distinct from intelligent machines even while the two are becoming increasingly entwined' (2005, p. 242). She suggests that if, even as they are drawn together, a clear idea of the remaining difference between human and machine is retained, the development of 'a dynamic partnership between humans and intelligent machines' can be proposed (Hayles, 1999, p. 288). (Sandry, 2015, p. 91)

We can understand what machines are and are becoming only if we can understand who we are. Technology is a key place to study humans because humans have always lived technological lives. However, the two concepts are both elusive. So far, instead of working hard on this issue, several scholars have preferred to indulge in a kind of praise of the machines and their intelligence and smartness. We try to initiate a path that should be able to help us answer these two questions in the future, although we are afraid, much time and effort will be needed to arrive at some substantial responses.

So, what is a robot, and what are its capabilities? Merely arriving at a definition of robot is difficult (Sarrica et al., 2020). Not by chance, Nourbakhsh (2013, pp. 14–15) wrote "never ask a roboticist what a robot is. The answers change too quickly." The robot body is too multiform and too mutant to allow a univocal answer to the extent that from a semiotic perspective, they can be considered a highly variable geometrical machine (Montanari, 1999). The multiform body of robots is given by the fact that robots may assume many forms: androids and gynoids, machine-like (such as Roomba, Bimby, Turtlebot), plantoids, zoomorphic and theomorphic robots, drones, voice-based assistants, intelligent agents, computer AI programs, chatbots, ambient assistive living technologies, computational, intelligent games/storytelling devices, embodied conversational avatars, automatic health care, and educational services. The database ABOVE available in open-source format (https://www.researchgate.net/publication/341835020_A_Comprehensive_Approach_to_Validating_the_Uncanny_Valley_using_the_Anthropomorphic_RoBOT_ABOT_Database [accessed Nov 08 2020]) includes 251 different types of real robots even if it is far from being exhaustive. Second, changes in robots' bodies are very fast and not always intelligible, as many of these kinds of robots are still at the prototype level. The prototyping status implies that we do not know many things we should do. Bhattacharya (2020), in his seminar, raised the example of GPT-3T, a language learning program able to talk in the style and manner of a famous psychologist, to conclude that we don't really know the linguistic capabilities of machine learning, because the machine may not react the way you think. The tremendous advances that are made in this kind of computer AI programs lead them to undertake very sophisticated interpersonal conversations and give responses that you cannot anticipate. Robots have such a multiform and mutant body that it becomes difficult to talk of robots' identity as well as of robots' capabilities, in general, because there are too many differences between them.

As to the other element of communication and interaction, the human being, we pose the same questions: What is a human being? What are their capabilities regarding thinking and doing things? In what ways are these capabilities different from those of robots? Reflections about what a human being is are first of all pertinent to the philosophical field and in fact some philosophers elaborated thoughts over time, although in a non-systematic way. Bhattacharya (2020) recalls some pivotal moments and protagonists of these reflections and focuses on the Western Europe of the XVII century. In this century, a series of wars (like the civil war between Protestants and Catholics, the Thirty Years War, and the development of the colonization of the Americas) bloodied the world, leading philosophers to ask themselves the question of what human beings were. Thomas Hobbes tried to answer the question by embracing the old Latin phrase *homo homini lupus* ("man is wolf to man") and describing human society as *bellum omnium contra omnes* ("the war of all against all"), when the situation that characterizes the conditions of the state of nature occurs. René Descartes contributed to the reflection on the definition of what humans are by arguing that the human being, like all animals, is a machine constituted by sub-machines. Strangely enough, comments Bhattacharya (2020), "modern medicine is still based on such a hypothesis." Michel de Montaigne, living in a society experiencing colonialism, faced the problems that in America Europeans encountered very strange kinds of people (e.g., cannibals). Were they human beings or not? The answer he found to this question was that the human being is a product of the culture in which they are raised. Montaigne stated, continues Bhattacharya (2020), that if he had been raised in a cannibalistic society, he too would have been a cannibal. Human beings are raised in cultural frameworks that shape their thinking and feeling. As Tomasello (1999/2005) argued, a human nature independent from culture does not exist; human nature is culturally situated to the extent that we should talk of *human natures*.

Why do we have cultures? We have cultures for the same reasons for which we have societies. According to Anolli (2005), cultures are collective constructions, which give shape and substance to human existence and define a specific perspective on reality. Tomasello (1999/2005) conceptualizes culture as a process and argues that humans are the outcomes of a double evolution: biological and cultural. Evolutionary social sciences so far have reconstructed humans' evolution or by contrasting them with animals and in particular with nonhuman primates or contrasting among them key types of social organizations, such as small-scale societies and large-scale societies (Seabright et al., 2021). In the first case, the comparison has always been made with nonhuman primates, because these show a little genetic difference (which, in the case of chimpanzees, is only 1%). What is unique to human beings, according to Tomasello (1999/2005), is a series of capabilities, which are made possible by humans' high propensity to sociability. They are the abilities to understand and interpret others as communicative and intentional agents; to build a network of bonds and motivations in the groups that are the basis of communication; to spontaneously share information and experience; to imitate the actions of others; to pursue the joint sharing of attention toward an object or event that implicates the focalization of the reciprocal psychological resources on the same object or event; to manage the dissociation between means and purposes, which, when practiced in pretend play, favors learning by simulation.

All of these capabilities are at the basis of cultural learning and lead to that symbolic condition, which for Deacon (1977) was the species-specific characteristic of humans. The great propensity for sociality that distinguishes human beings, also nourishes the processes of sociogenesis, which, according to Tomasello (1999/2005), thanks to the social cooperation of a multiplicity of individuals, are effective forms of creation and invention. These creative forms are at the base, he continues, of language and mathematics, the majority of cultural practices and artifacts, and cultural learning. All of these processes, argues

Tomasello, have transformed some phenomena, which were typical of all the primates, such as communication, dominance, exchange, and explorative behavior, in human cultural and social institutions like language, state, money, science.

In the second case, evolutionary social scientists such as Seabright et al. (2021) looking at small-scale societies offer another series of interesting findings for our topic of interest here. They document that in this type of societies humans have an evolved cognitive specialization for reasoning about social exchange and social learning. Furthermore, in small-scale societies, continue Seabright et al. (2021), humans are used to several enjoyable activities, such as storytelling, music-making, singing and dance, sport, and communal beer-drinking. These activities, which have been handed down over time to the present day, are those we call today "forms of communication sociability" (Fortunati et al., 2013) and that play a central role in socialization, information exchange, and/or entertainment. Furthermore, in small societies, ethical behaviors are appreciated because they help the social organization. Saucier et al. (2014), for example, demonstrate that a lexical study of traits ascribable to humans in 12 isolated languages covering most habitable world regions outside of Europe show jealousy and crookedness are relatively ubiquitous human traits. Purzycki et al. (2018) similarly identified honesty and dishonesty as prominent cross-cultural indicators of good and bad people, respectively. In addition to such behavioral norms, there is widespread acceptance of the idea that one should aspire to find virtuous people as social partners, and avoid anti-social individuals. Moreover, cooperation and collaboration are very much appreciated by humans in small-scale societies whereby cooperators prefer to connect to other cooperators in social networks (Smith et al., 2017; Stieglitz et al., 2017). These collaborative activities include both those that are meant to provide public goods and those that can be described as rituals, in which the activity itself represents the collective benefit being provided. Bhattacharya (2020) provides another key insight: Most of what we are is biopsychological and our selves are mostly settled on the basis of upbringing. These two elements are evolutionary products that shape our lives. Sketched above what a human is, we can initiate tracing some meaningful dividing lines between a present-day robot and human. Of course, we are not able to make a precise and exhaustive analysis and, as we noted above, both the nature of robots and concepts of humanity shift over time, but we can at least outline some elements of remaining and important distinction.

- A robot is not born from a mother and cannot experience pregnancy and giving birth.
- A robot has no infancy and is not shaped by a community of adults (experiencing primary and secondary socialization). In fact according to Gray and colleagues (2007), whereas humans are perceived to be high in agency and experience, animals like dogs are considered high in experience and low in agency, and robots are perceived to be high in agency and low in experience.
- A robot does not have consciousness of itself and its being in the world (Faggin, 2021).
- A robot does not feel emotions, amuse itself, and pretend to care.
- A robot is not a cultural subject, only a cultural object.
- A robot does not share food and drinks with humans.

- A robot does not have an authentic comprehension of social institutions.
- A robot does not have goals or purposes, as Bhattacharya (2020) points out, "beyond carrying out a pre-assigned program."
- A robot has no conception of what humans do and why.
- A robot does not perceive the environment in sociocultural, emotional, and spiritual ways, only in a physical way.

The drawing of boundaries around and between humans and machines (i.e., ontologizing) is closely related to the process of prototyping, or determining what something exemplifies by fitting it within an overarching category. Especially when we encounter new interlocutors, relationship configurations, and interaction contexts, as is often the case in HMC, we ask ourselves, "What *is* it?" As the constructivist communication tradition has long stressed, prototyping is important because our answers to that question will lead us to activate certain stereotypes (what does it *do*?) which will then influence perception and communication by guiding the application of mental constructs and the activation, modification, and use of interaction scripts (Edwards, 2018, p. 46; see also Edwards et al., 2019). In their article, "Forms and Frames: Mind, Morality, and Trust in Robots Across Prototypical Interactions," Banks et al. advance the knowledge on prototyping in HMC by focusing on the prototypicality of interaction, or the degree to which a human-robot interaction represents its larger interaction form. Banks et al. propose social, task, and play as three distinct higher-order interaction forms and examine via experimental methods their differential effects on how people perceive the mind, morality, and trust (worthiness) of a humanoid social robot. Although mind ascription and moral evaluations did not differ based on the interaction form, participants expressed more trust following the task interaction, a form of interaction the authors suggest may be most aligned with our common mental models regarding the nature, behavior, and skills of robots. In the end, participants' *perceptions* of the interactions were more influential to their trust and attitudes than were formal properties of the interaction, leading Banks et al. to conclude that "schema-aligned task interactions and perceived playfulness are gateways for building trust in robots" (p. 97). Trust remains a crucial issue in HMC. As Lutz and Tamó-Larrieux (2020) argued in Volume 1 of *Human-Machine Communication* ("The Robot Privacy Paradox: Understanding How Privacy Concerns Shape Intentions to Use Social Robots") trust is a key element for developing comfort with social robots and determining their appropriate and ethical uses.

## 3. Overcoming Binary Discourse: A Critical Turn in Human-Machine Communication Research

The third major strand in this volume includes three articles that challenge dominant but unquestioned discourses of power and privilege infusing HMC. In the article "Social Robots as the Bride? Understanding the Construction of Gender in a Japanese Social Robot Product" Liu concentrates on Azuma Hikari, the beautiful anime-style girl that is produced as a hologram device by Gatebox Inc. As opposed to voice-based assistants like Siri, Alexa, and Cortana, Azuma Hikari is designed as an intimate companion, a "comforting bride" with whom to simulate marriage. Using a mixed-methods approach, Liu combines semiotic analysis of the official visual materials released by Gatebox Inc. with the heterogeneous engineering approach proposed by Gehl (2014) to critically consider the obscured and hidden aspects of Azuma Hikari that in daily interactions with users become salient alongside the official marketing materials. Liu's analysis reveals that Azuma Hikari leverages stereotypical representations of a Japanese "ideal bride" and reproduces through interaction an unequal gender relation where users are disciplined to play the role of "wage earner" and master-like husband. As Liu concludes, the humanization of objects is often associated with the objectification of humans. This research is a valuable contribution that encourages systematic attention to how HMC is reflective and productive of human gender relations and also models an exemplary methodological approach for doing this work.

Davis and Stanovsek, in their article "The Machine as an Extension of the Body: When Identity, Immersion, and Interactive Design Serve as Both Resource and Limitation for the Disabled," focus on the interplay between the technological affordances of a social virtual world and the people with disabilities (PWD) who interact via avatars in this online community. Arising from their ongoing, multiyear ethnographic study of embodiment among PWD in the virtual world *Second Life* (SL), Davis and Stanovsek address two research questions: (1) How do PWD who function as avatars report the technological affordances of virtual environments as functions of online identity in support of social or professional interaction? and (2) What elements of platform design (affordances) either support or inhibit social function among PWD in SL?

Results from thematic analysis demonstrate the important relationship between avatar choice, which may be congruent or incongruent with the user's physical self along many identity markers, and the development of identity and also the fundamental and increasing integration of technology itself with identity, ability, and access to professional and social communities. Through their analysis, Davis and Stanovsek further show that "while these affordances may allow for an expression of identity tied to ability, creativity and sociability not accessible in the physical world," they also "create new barriers to accessibility for individuals who have found freedoms in their embodied online surrogate" (p. 133).

Dehnert and Leach in their article "Becoming Human? Ableism and Control in *Detroit: Become Human* and the Implications for Human-Machine Communication," present a case study of the choose-your-own-adventure video game *Detroit: Become Human* to reveal how people may rely on ableist discourses to make sense of relationships with machines. They detail how ableist communication scripts cast machine partners as both sub- and superhuman, signifying lack and also excess in relation to the socially constructed "normal body," and how those scripts manifest in control and cyborg anxiety (i.e., "the fear that the performance and embodiment of a disabled person exposes the porous and permeable boundaries of what it means to be an able-bodied and able-minded human," p. 139). This case study serves as a demonstration and vehicle for Dehnert and Leach's broader call for critical approaches to understanding the communication scripts people use in HMC. In their words:

> A critical perspective on a constructivist research paradigm in HMC attunes us to the ubiquitous yet powerful societal systems of oppression that guide communication scripts: structures such as heteronormativity and Whiteness that have been intensively theorized by critical communication scholars. (p. 138)

Rightly, Dehnert and Leach contend that uncritical acceptance of the automatic (naïve, mindless) transfer of scripts developed in human interaction will thwart our ability to resist and rescript harmful relations, both human and human-machine.

These articles regarding gender (man and woman) and body representation (ability and disability) reveal the necessity to resume the debate that developed in the 1980s and that advocated a radical overcoming of binary thinking. In those years the feminist movement contested a knowledge resulting from an abstract form of the real, based on dialectic and structured with such characteristics to preclude women, by forbidding them access to thought and, circularly, by prohibiting thought from accessing the world of women (e.g., Bordo, 1986; Costantini et al., 1981; Harding & Hintikka, 1983; Janssen-Jurreit, 1982; O'Brien, 1981). Negri (1981) also wrote about dialectic in recognizing the rupture of this eternal formula of Judeo-Christian thought within political discourse. But the first contestation of Western thought, the disavowal of its formulas and its categories, starting with dialectic, dualism, and beyond which everything that was there was branded with a generic label of irrational—from circumstantial knowledge to emotionality, from bodily knowledge to participatory intuition—was put in place by women. The contestation of the binary discourse was necessary for women. The feminist movement not only criticized how gender and identity had been shaped but at the same time defended all the subjects who were beyond the famous fixed tracks. Today we welcome the notion of "gender continuum," which includes third genders, agender, two-spirit, and different dimensional models of gender (Søraa, 2017), but this has been made possible by the rupture of the binary perspective.

There is the same need to overcome the binary discourse involved in the dialectic abled/disabled, which is still too often based on the stereotyped, social representation of the abled as normal and the point of reference and the disabled as abnormal and the exception (Davis, and Denhert & Leach in this volume). This vision is not only simplistic, narrow, and stigma-generating, but it also prevents us from seeing the reality of the so-called normal, which is imbued with disability during the life cycle. When able people become elderly they face and live with many disabilities: for example, they become visually- and hearing-impaired, they may have problems walking or have to move in a wheelchair, and so on (Cortés et al., 2003). Furthermore, chronic diseases are significant causes of disability and reduced quality of life for able people. Infancy is also a period in which children face many types of disability and need a long time to reach their muscle development and control mastery. Ability is something that children need to conquer through movement. Thus, disabled people are a very heterogeneous group and include as we mentioned above some who are "learning to be able" (children) and ex-able social groups (elderly), whereby scientific literature has increasingly begun to study them together with elderly people (e.g. Lee et al., 2021; Madhusanka & Ramadass, 2021). The history of digital technologies has documented that the devices designed for disabled people have been a frontier of technological inclusion since they have allowed the design of better technologies for all: see the debate on universal design versus assistive technologies (Vanderheiden, 1998) and the advantages and disadvantages related to each of these approaches.

## 4. Exploration of Human-Machine Communication in Organizational Contexts

A final group of four papers focuses on the exploration of HMC in different sectors of production. Gibbs et al., in their article "Negotiating Agency and Control: Theorizing Human-Machine Communication From a Structurational Perspective," draw on structuration theory to propose a conceptual framework to address HMC in organizational contexts. They note that in contrast with earlier organization technologies theorized primarily as tools, intelligent technologies may be considered agents with the potential to transform organizing and organizations. However, a robust understanding of the role of smart and communicative machines in organizational communication will require moving beyond individual-level analyses of agency and control. Gibbs et al. forward structuration theory (Giddens, 1984) as a means of recognizing the recursive interplay of structure and agents in which each constitutes the other; agents create structures which both enable and constrain their practices, and their practices reproduce social systems and structures. They offer a theoretical framework to explain agency in HMC as "a process involving the negotiation of control between human and machine agents" (p. 153). Through illustrative organizational examples of A/B testing, humans in the loop, and automated journalism, Gibbs et al. demonstrate the value of the *dialectic of control*; the "reciprocal power relationship between active agents and institutional forms of control" (p. 161) in organizational contexts which now include algorithmic as well as human and institutional forms of control. They conclude their discourse by presenting an exciting and important future research agenda for the application of the structurational framework in the HMC field.

Johanssen and Wang, in "Artificial Intuition in Tech Journalism on AI: Imagining the Human Subject," explore the field of tech journalism by presenting an interesting analysis of journalistic frames and discussing the new trend of Artificial Intuition (AI acting intuitively). Through a qualitative thematic analysis of tech journalism articles (from 2016–2019) that discussed Artificial Intuition as it relates to human subjectivity, Johanssen and Wang demonstrate that the narratives typically claimed intuition could make AI more efficient, autonomous, and human, and sometimes claimed that it could outperform or surpass human abilities and intuition. The rational/computation framing of tech journalism is examined in contrast to philosophical and psychoanalytic explanations of intuition, which introduce greater complexity and call for more attention to explainability, transparency, and bias in AI systems.

Piercy and Gist-Mackey in "Automation Anxieties: Perceptions About Technological Automation and the Future of Pharmacy Work" investigate workers' anxiety in respect to the diffusion of automation in the pharmacy context. Noting an increasingly technologizing work landscape, they orient their research toward the general question: "How do employees perceive the future of human-technology collaboration?" (p. 192) Specifically, they test the *skills-biased technical change hypothesis (SBTC hypothesis)* which predicts that low-skill work is the most likely to be transformed or replaced in automation processes by examining the mental models about automation in a sample of pharmacists and pharmacy technicians. Their results are largely supportive of the SBTC hypothesis and demonstrate the importance of examining the influence of anxieties on mental models surrounding HMC issues.

Prahl and Van Swol in their article "Out With the Humans, in With the Machines? Investigating the Behavioral and Psychological Effects of Replacing Human Advisors With a Machine" analyze the behavioral and psychological effects of replacing humans with robots in the financial sector. The authors report the results of an experiment in which participants were tasked with making a series of financial forecasts in a low or high demonstrability situation after receiving advice from a machine advisor, a human advisor, or an advisor replaced by the other agent type (human-machine) halfway through the experiment. Human advisors were rated as broadly more favorable (expert, useful, similar), an effect magnified when the human was replaced by a machine during the experiment. Prahl and Van Swol discuss the numerous and complex factors that must be considered when effectively introducing machine advisors into organizational contexts and note that "it is not only humans who are replaced with will be unhappy; the people who work with these new machines may not be happy either" (p. 226).

All of these articles address timely and important questions related to various organizational contexts and contribute to moving the research on HMC ahead. They also indicate the need to acknowledge that the penetration of robotics, virtual assistants, intelligent agents, chatbots, and so on in the production sphere takes place along with the penetration and development of a cluster of processes or factors as indicated by Evans (2017): soft automation, digitalization, AI, big data, social media, and 3D printers. These factors interpenetrate and empower each other creating smart and networked environments both in factories and services and in the urban setting. Together, they are reshaping the industrial world in the direction of cutting production costs by increasingly reducing the use of the human labor force. However, as Horkheimer and Adorno pointed out in 1966,

> what the modern-day technical civilization has of chaotic and monstrous does not derive from the very idea of a technical civilization or from some essence of technology as such. In modern society, technology has acquired by now a characteristic position and structure, whose relationship with the needs of human beings is profoundly incongruous: evil, therefore, does not derive from the rationalization of our world, but from the irrationality with which that rationalization takes place. (p. 108)

It is the perception and the experience of this irrationality that create, for example, the pharmacy workers' anxiety which Piercy and Gist-Mackey investigate or the negative emotions and mistaken behavior that decision-makers experience when a human is replaced by a machine in the financial sector, documented by Prahl and Van Swol.

Although the research on HMC in organizational contexts we publish in this volume is interesting and relevant, there is also the need for innovative research on self-driving cars. After the special issue of the *International Journal of Communication* (vol. 13, 2019) "Cars and Contemporary Communication: Machine, Medium, Mobility" edited by Thilo von Pape, Gerard Goggin, and Laura Forlano, it is time to advance the debate on communication between humans and this type of machine. This debate does not involve only the self-driving cars, the cars of the future, but also and especially the automation processes involved in the cars that we drive today and the cars that we will buy tomorrow. Important processes of robotization, as argued by Fortunati and colleagues (2019), concern current cars as well as many other machines that populate our everyday life. These processes have taken place against a backdrop in which cars have become over time a central hub of social, communication, and technology developments. Today, cars are a key place of encounter and hybridization between mobility, media use, mediated and analog communication, and activities such as driving. Furthermore, they have been transformed into a complex node of communication flows with different degrees of automation. Drivers are bombarded by flows arriving from road infrastructures such as traffic lights and road signs, from the car itself such as satnav systems and geowebs supporting driving and wayfinding, from pedestrians, as well as from car passengers. Drivers in turn have to communicate with pedestrians directly and through the car, with other cars, the road infrastructure, and the other passengers. On top of these information exchanges, Goggin (2012) argued that in-car communication also flows across multiple channels: mobile phones, internet, TV, and radio, with the consequence that drivers assume many different communicative roles: they are conversation interlocutors, users of mobile media, and audiences. It is extremely important for the HMC field and for humans that we advance the study of robotization processes in this communication context.

## 5. Conclusion

The four layers of discourse that this second volume of *Human-Machine Communication* contains substantially advance key debates in the field. The foundations of the major theoretical frameworks are progressing and the methodological tools are subject to ongoing innovation. Furthermore, the research methodologies used in this field are less confined to strictly quantitative and experimental ones, and increasingly integrate qualitative and mixed methods. Likewise, it is encouraging to witness critical approaches to HMC emerging alongside and also challenging the more traditional and administrative approaches to theory and practice. We still need to better understand the actual practices of use of technologies, which are at the basis of the big issues, if we want to progress substantially in this field (Suchman, 2019). However, we are aware that we face serious limitations in exploring the daily practices of use, especially of social robots, because these are still at the prototyping level and, thus, users have a very limited direct experience of them. This call for additional research on situated use practices, as well as our call for greater integration of historical, sociological, and semiotic/hermeneutic insights, for work that interrogates and challenges structures of privilege, exploitation, and oppression, and for research that extends HMC inquiry into transforming and transformative communication technologies like the automobile, is offered with excitement and appreciation for our wonderful community of researchers, theorists, and practitioners. We extend our gratitude to the brilliant, generous, and constructive scholars who comprise the editorial board and to the scientists, critics, and practitioners who submitted their excellent work.

## References


Anolli, L. (2005). Introduction. In M. Tomasello. Le origini culturali della cognizione umana (pp. 7–18). Bologna: il Mulino.

Appadurai, A. (Ed.) (1986). The social life of things: Commodities in cultural perspective. Cambridge University Press.

Baron, N. (2013). Authenticity, emotions, and ICTs. *Intervalla: Platform for Intellectual Exchange*, *1*, 7–16.

Barthes, R. (1957). Mythologies. Éditions du Seuil.

Bhattacharya, N. (2020). "The human and the robotic: Mutual interaction," seminar held at the University of Udine the 1st December 2020.

Bordo, S. (1986). The Cartesian masculinization of thought. *Signs: Journal of Women in Culture and Society*, *11*(3), 439–456. https://doi.org/10.1086/494250

Cecchi, D. (2019). Il lettore esemplare: Fenomenologia della lettura ed estetica dell'interazione. Rivista di estetica 71: 257–270.

Chatman, S. B. (1980). *Story and discourse: Narrative structure in fiction and film*. Cornell University Press.

Cometa, M. (2017). *Perché le storie ci aiutano a vivere*. Milano: Raffaello Cortina.

Consoli, G. (2011). L'immaginazione al lavoro. Che cosa e come conosciamo dal romanzo. [Imagination at work. What and how do we know from the novel]. ENTHYMEMA (5), 102–116. https://doi.org/10.13130/2037-2426/1756

Cortés, U., Annicchiarico, R., Vázquez-Salceda, J., Urdiales, C., Cañamero, L., López, M., Sànchez-Marrè, M., & Caltagirone, C. (2003). Assistive technologies for the disabled and for the new generation of senior citizens: The e-Tools architecture. *AI Communications*, *16*, 193–207.

Cosenza, G. (2009). Semiotica dei nuovi media. [Semiotics of New Media]. *Enciclopedia Treccani*. https://www.treccani.it/enciclopedia/semiotica-dei-nuovi-media_%28XXI-Secolo%29/

Costantini, S., Gagliasso, E., & Tagliaferri, M. (1981). Il luogo delle ipotesi. *Memoria* 1: 9–27.

Danaher, J. (2017). The symbolic-consequences argument in the sex robot debate. In J. Danaher & N. McArthur (Eds.) *Robot sex: Social and ethical implications*. MIT Press.

Deacon, T. W. (1977). *The symbolic species. The co-evolution of language and the human brain*. Norton.

Dehaene, S. (2007). *Les Neurones de la lecture,* Paris, Odile Jacob, trad. it. C. Sinigaglia, I neuroni della lettura, Milano, Raffaello Cortina 2009.

Douglas M., & Isherwood B. (1979). *The world of goods. Towards an anthropology of Consumption*. Basic Books.

Eco, U. (1983). *Lector in fabula. La cooperazione interpretativa nei testi narrativi*. Bompiani.

Eco, U. (1985). *Sugli specchi e altri saggi*. Bompiani.

Eco, U. (2016). *Sei passeggiate nei boschi narrativi*. Bompiani.

Edwards, A. (2018). Animals, humans, and machines: Interactive implications of ontological classification. In A. Guzman (Ed.) *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 29–49). Peter Lang.

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior, 90*, 308–314. https://doi.org/10.1016/j.chb.2018.08.042

Evans, S. (2017). On the efficiency of our industrial sector system as a stakeholder in our natural system. Presentation at the annual conference of the Academia Europaea "Sustainability and Resilience," Budapest, 3–5 September 2017.

Faggin, F. (2021). *Silicon: From the invention of the microprocessor to the new science of consciousness* [Audiobook]. Waterside Production, Inc.

Fehr, E., & Fischbacher, U. (2004). Third party punishment and social norms. *Evolution and Human Behavior*, *25*(2), 63–87. https://doi.org/10.1016/S1090-5138(04)00005-4

Fischbacher, U. (2001). Are people conditionally cooperative? *Economics Letters*, *71*(3), 397–404. https://doi.org/10.1016/S0165-1765(01)00394-9

Fortunati L., Lugano G., & Manganelli A. M. (2019). European perceptions of autonomous and robotized cars. *International Journal of Communication*, *13*(2), 2728–2747.

Fortunati, L., Taipale, S., & De Luca, F. (2013). What did happen to body-to-body sociability? *Social Science Research*, *42*(3), 893–905. https://doi.org/10.1016/j.ssresearch.2012.12.006

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 71–86. https://doi.org/10.30658/hmc.1.5

Gehl, R. (2014). *Reverse engineering social media: Software, culture, and political economy in new media capitalism*. Temple University Press.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.

Goggin, G. (2012). Driving the Internet: Mobile internets, cars, and the social. *Future Internet*, *4*(1), 306–321. https://doi.org/10.3390/fi4010306

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*, 619. https://doi.org/10.1126/science.1134475

Grier, D. A. (2005). *When computers were human*. Princeton University Press.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, *1*, 37–54. https://doi.org/10.30658/hmc.1.3

Harari, Y. N. (2016). *Sapiens*. Bazarforlag AS.

Harding, S., & Hintikka, M. (1983). *Discovering reality: Feminist perspectives on epistemology, metaphysics, methodology, and the philosophy of science.* Reidel Publ.

Hayles, N. K. (1999). *How we became posthuman: Virtual bodies in cybernetics*. University of Chicago Press.

Hayles, N. K. (2005). *My mother was a computer: Digital subjects and literary texts*. University of Chicago Press.

Hicks, M. (2018). *Programmed inequality: How Britain discarded women technologists and lost its edge in computing*. MIT Press.

Horkheimer, M., & Adorno, T. W. (1966). Dialettica dell'illuminismo. Einaudi, Torino 1966. [Dialectic of Enlightenment. Philosophical Fragments. Stanford University Press, 2002].

Iser, W. (1976/1987). *Der Akt des Lesens*. München: Fink (Trad. It. L'atto della lettura. Una teoria della risposta estetica. Bologna: il Mulino).

Janssen-Jurreit, M. L. (1982). *Sexism: The male monopoly of history and thought*. Strauss & Giroux.

Jauss, H. R. (1967/1969). *Literaturgeschichte als Provokation der Literaturwissenschaft*. Konstanz: Universitäts-Druckerei [Trad. It. Perché la storia della letteratura? Napoli: Guida.

Knorr Cetina, K. (1997). Sociality with objects. Social relations in postsocial knowledge societies. *Theory, Culture & Society*, *14*(4), 1–30. https://doi.org/10.1177/026327697014004001

Latour, B. (1994). Une sociologie sans objet? Note théorique sur l'interobjectivité, Sociologie du travail, vol. XXXIV(4), pp. 587–607 (It. Transl. Una sociologia senza oggetto? Note sull'interoggettività. In E. Landowski and G. Marrone (eds.) *La società degli oggetti. Problemi di interoggetività*, Roma: Meltemi. pp. 203–232).

Lee, J. M., Kim, Y. M., Rhiu, I., & Yun, M. H. (2021). A persona-based approach for identifying accessibility issues in elderly and disabled users' interaction with home appliances. *Applied Sciences*, *11*, 368. https://doi.org/10.3390/app11010368

Lutz, C., & Tamò-Larrieux, A. (2020). The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots. *Human-Machine Communication*, *1*, 87–111. https://doi.org/10.30658/hmc.1.6

Madhusanka, B. G. D. A., & Ramadass, S. (2021). Implicit intention communication for activities of daily living of elder/disabled people to improve well-being. In G. Marques, A. K. Bhoi, V. H. C. de Albuquerque, & K. S. Hareesha (Eds.), *IoT in Healthcare and Ambient Assisted Living* (pp. 325–342). *Studies in Computational Intelligence* 933, Springer.

Montanari, F. (1999). Dall'oggetto al fatticcio. [From the object of the story]. In G. Marrone (Ed.) *C'era una volta il telefono (pp. 169–192). Roma: Meltemi.*

Montani, P. (2014). *Tecnologie della sensibilità*. Raffaello Cortina.

Mori, M. (1970). The uncanny valley. *Energy* 7(49): 33–35. (K. F. MacDorman & T. Minato, Trans.). http://www.movingimages.info/digitalmedia/wp-content/uploads/2010/06/MorUnc.pdf

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human Computer Studies*, *45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

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., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). ACM. https://doi.org/10.1145/259963.260288

Nass, C. I., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship.* MIT Press.

Negri, A. (1981). Erkenntnistheorie. Elogio dell'assenza di memoria. *Metropoli 5:* 50–52.

Nourbakhsh, I. R. (2013). *Robot futures*. Mit Press.

O'Brien, M. (1981). Feminist theory and dialectical logic. *Signs: Journal of Women in Culture and Society*, *7*(1), 144–157.

Ogden, C. K., & Richards, I. A. (1923). *The meaning of meaning*. Kegan Paul.

Ong, W. J. (2002/1986). *Orality and literacy: The technologizing of the word*. Routledge [Trad. It. Oralità e scrittura. Le tecnologie della parola, Bologna: Il Mulino].

Peirce Edition Project, & Peirce, C. S. (1998). *The Essential Peirce, Volume 2*. Indiana University Press.

Pietschnig, J., & Voracek, M. (2015). One century of global IQ gains: A formal meta-analysis of the Flynn Effect (1909–2013). *Perspectives on Psychological Science, 10*(3), 282–306. https://doi.org/10.1177/1745691615577701

Purzycki, B., Pisor, A., Apicella, C., Atkinson, Q., Cohen, E., Henrich, J., McElreath, R., McNamara, R., Norenzayan, A., Willard, A., & Xygalatas, D. (2018). The cognitive and cultural foundations of moral behavior. *Evolution and Human Behavior, 39*, 490–501. https://doi.org/10.1016/j.evolhumbehav.2018.04.004

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

Ricœur, P. (1983–1985/1986–1988). Temps et récit, 3 voll. Paris: Gallimard. [Trad.It. Tempo e racconto, 3 voll. Milano: Jaca Book].

Rizzolatti, G., & Sinigaglia, C. (2006). *So quel che fai*. Milano: Raffaello Cortina.

Saintato, M. (2020, Feb 5). 'I am not a robot': Amazon workers condemn unsafe, grueling conditions at warehouse. *The Guardian*. Retrieved March 15, 2021, from https://bit.ly/30Od8ye

Sandry, E. (2015). *Robots and communication*. Palgrave.

Sarrica, M., Brondi, S., & Fortunati, L. (2020). How many facets does a "social robot" have? A review of scientific and popular definitions online. *Information Technology & People, 33*(1), 1–21. https://doi.org/10.1108/ITP-04-2018-0203

Sartre, J.-P. (1948/1960). *Qu'est-ce que la littérature ?* Paris: Gallimard [trad. it. Che cos'è la letteratura? Milano: Il Saggiatore].

Saucier, G., Thalmayer, A. G., & Bel-Bahar, T. S. (2014). Human attribute concepts: Relative ubiquity across twelve mutually isolated languages. *Journal of Personality and Social Psychology*, *107*(1), 199–216.

Seabright, P., Stieglitz, J., & Van der Straeten, K. (2021). Evaluating social contract theory in the light of evolutionary social science. *Evolutionary Social Sciences*, online first, pp. 1–45. https://doi.org/10.1017/ehs.2021.4

Shneiderman, B., Plaisant, C., Cohen, M. S., Jacobs, S. A., & Elmqvist, N. (2017). *Designing the user interface: Strategies for effective human-computer interaction* (6th ed.). Pearson

Simmel, G. (1903/1995). *Die Großstädte und das Geistesleben* [trad.ingl. The Metropoles and the Life of Spirit; trad.it. La metropoli e la vita dello spirito, edited by P. Jedlowski: Roma, Armando Editore, 1995].

Smith, D., Schlaepfer, P., Major, K., Dyble, M., Page, A., Thompson, J., Chaudhary, N., Salali, G., Mace, R., Astete, L., Ngales, M., Vinicius, L., & Migliano, A. (2017). Cooperation and the evolution of hunter-gatherer storytelling. *Nature Communications*, *8*, 1853.

Sneddon, A. (2015). Symbolic value. *Journal of Value Inquiry*, *50*(2), 395–413. https://doi.org/10.1007/S10790-015-9519-4

Søraa, R. A. (2017). Mechanical genders: How do humans gender robots? *Gender, Technology and Development, 21*(1–2), 99–115. https://doi.org/10.1080/09718524.2017.1385320

Stieglitz, J., Gurven, M., Kaplan, H., & Hopfensitz, A. (2017). Why household inefficiency? An experimental approach to assess spousal resource distribution preferences in a subsistence population undergoing socioeconomic change. *Evolution and Human Behavior*, *38*, 71–81. https://doi.org/10.1016/j.evolhumbehav.2016.07.002

Suchman, L. (2019). If you want to understand the big issues, you need to understand the everyday practices that constitute them: Lucy Suchman in Conversation With Dominik Gerst & Hannes Krämer. Forum: *Qualitative Social Research*, *20*(1). https://doi.org/10.17169/fqs-20.2.3252

Tomasello, M. (1999/2005) *The cultural origins of human cognition*. Harvard University Press [Trad. it. Le origini culturali della cognizione umana. Bologna: il Mulino].

Vanderheiden, G. (1998). Universal design and assistive technology in communication and information technologies: Alternatives or complements? *Assistive Technology: The Official Journal of RESNA*, *10*(1), 29–36. https://doi.org/10.1080/10400435.1998.10131958

von Pape, T., Goggin, G., & Forlano, L. (Eds.) (2019). Cars and contemporary communication: Machine, medium, mobility. *International Journal of Communication*, *13*, 2676–2683.

Wolf, M. (2008/2009). *Proust and the squid*. Harper Perennial [Trad. It. Proust e il calamaro. Milano: V&P].

Human-Machine Communication
Volume 2, 2021
https://doi.org/10.30658/hmc.2.2

# Social Responses to Media Technologies in the 21st Century: The Media Are Social Actors Paradigm

**Matthew Lombard**[1] **and Kun Xu**[2]

1 Klein College of Media and Communication, Temple University, Philadelphia, PA, USA
2 College of Journalism and Communications, University of Florida, Gainesville, FL, USA

## Abstract

Clifford Nass and his colleagues proposed the Computers Are Social Actors (CASA) paradigm in the 1990s and demonstrated that we treat computers in some of the ways we treat humans. To account for technological advances and to refine explanations for CASA results, this paper proposes the Media Are Social Actors (MASA) paradigm. We begin by distinguishing the roles of primary and secondary cues in evoking medium-as-social-actor presence and social responses. We then discuss the roles of individual differences and contextual factors in these responses and identify mindless and mindful anthropomorphism as two major complementary mechanisms for understanding MASA phenomena. Based on evolutionary psychology explanations for socialness, we conclude with nine formal propositions and suggestions for future research to test and apply MASA.

**Keywords:** computers are social actors, media are social actors, medium-as-social-actor presence, social presence, social cues, mindlessness, anthropomorphism

## Introduction

Among the more surprising research results in the study of human-computer interaction are those of Clifford Nass and his colleagues demonstrating that computer users apply social rules from human-human interactions when they use computers. A series of experiments demonstrated that users follow social rules based on gender, team membership, politeness,

**CONTACT** Kun Xu • College of Journalism and Communications • University of Florida • 1885 Stadium Rd, Gainesville, FL 32611 • kun.xu@ufl.edu

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

and more (Nass et al., 1994; Nass et al., 1996; Nass et al., 1997). Nass proposed the Computers are Social Actors (CASA) paradigm to describe the phenomenon and an approach to studying it and expanded on the idea with Byron Reeves in their book *The Media Equation* (Nass et al., 1994; Reeves & Nass, 1996).

The original CASA paradigm was proposed over a quarter century ago. Today we are surrounded by many more media technologies, including wearable devices, smartphones, digital assistants, and humanoid robots (Carolus et al., 2019; Kanda et al., 2009; Perry, 2014). While growing research has been applying the CASA paradigm to understand interactions with these emerging technologies, the general tenet that users respond to computers and other media technologies as if they were real people requires more subtle and detailed explications that account for media users' behavior. Although the CASA paradigm and the Media Equation describe a series of experimental findings about individuals' social reactions to technologies, one caveat to the theoretical framework is that it lacks explicit propositions that researchers can examine, test, and refine. Thus, to build a more refined framework, we propose the Media are Social Actors (MASA) paradigm as a structured extension of the CASA paradigm. We suggest that an enhanced framework that builds on the CASA paradigm, expounds the effects of social cues, describes the psychological mechanism of social responses, and provides propositions that scholars can test will not only direct future research on human-technology interaction but also meet the criteria of explanatory power, predictive power, falsifiability, heuristic value, and internal consistency that Chaffee and Berger (1987) outlined for a rigorous theoretical framework.

We expand the CASA paradigm from the perspective of social cues, as social cues are a key factor that leads to users' social responses. For example, Nass (2004) noted that social cues which may trigger applying etiquette rules to computers include language, voice, a face, emotion manifestation, interactivity, perceived engagement with the user, and filling of social roles. Another reason to approach the expansion of the paradigm through social cues is that cues can evoke medium-as-social-actor presence. Medium-as-social-actor presence refers to the idea that when a medium itself presents social cues, individuals perceive it (with or without conscious awareness) not as a medium but as an independent social entity (Lombard & Ditton, 1997). While sometimes researchers use social presence to describe medium-as-social-actor presence experiences, social presence may involve other meanings such as the "subjective quality of the medium" (Short et al., 1976, p. 66) or parasocial interaction (Horton & Wohl, 1956). Thus, in this paper, we will consistently use the term medium-as-social-actor presence to avoid misunderstandings (for definitions of other types of social presence, see Biocca et al., 2003; Lombard & Jones, 2015). Although not all CASA studies have examined medium-as-social-actor presence, and presence need not always lead to behavioral social responses, social responses can be considered as a reflection and indication of users' medium-as-social-actor presence experiences.

While much literature has examined the effects of social cues on users' social responses to technologies (e.g., Araujo, 2018; Nomura & Kanda, 2015; Terzioğlu et al., 2020), most research has focused on either the effects of single social cues or the cumulative effects of social cues. A large body of research suggests that technologies with more social cues can evoke stronger social responses than ones with fewer social cues (e.g., Burgoon et al., 2000; Ghazali et al., 2018; Tung & Deng, 2007). By contrast, limited research has clarified the distinctions among these social cues. It is likely that individual social cues can exert different effects on users' responses (Abubshait & Wiese, 2017). Therefore, in this paper, we suggest that (1) the quality of social cues should receive more attention in future research, (2) the roles of individual differences and communication contexts should be considered in the paradigm, (3) the two prominent psychological explanations for social responses, mindlessness and anthropomorphism, can be unified and applied to explain a wide variety of human-machine communication (HMC) scenarios, and (4) specific propositions should be developed and refined to derive future research questions on HMC.

Expanding the CASA paradigm to the MASA paradigm through the examination of social cues has three major benefits. First, from a theory construction perspective, broadening the CASA paradigm will expand the value and applicability of an important theoretical framework. By refining and clarifying the concepts, assumptions, and propositions of the MASA paradigm, researchers can better understand, explain, predict the effects of, and generate new questions about our increasingly common communication experiences with diverse technologies.

Second, when interaction with technologies mimics natural human communication, people expend less cognitive effort (Gambino et al., 2020), making the experience more efficient and intuitive. And when interactions account for users' personalities, identities, and demographic characteristics along with the communication context, the experience is further enhanced. We can therefore apply the propositions of the MASA paradigm to improve the design, enhance the users' experience, and therefore increase the success, of a variety of media technologies (Bartneck et al., 2008; Biocca et al., 2003).

Third, expanding CASA to MASA highlights important ethical issues in technology development. As designers can more easily embed different social cues in technologies, we will need to consider the potential for unethical manipulation of user responses. Those who devise, promote, adopt, and study presence-evoking technologies should be mindful of their concomitant risks (Lombard, 2009).

## The Roles of Cues, Individual Differences, and Contextual Factors

To elaborate the mechanism of the MASA paradigm, below we first explicate the role of social cues. We demonstrate the intrinsic distinctions among social cues, the effects of social cues on medium-as-social-actor presence and social responses, and then briefly introduce the role of individual differences and contextual factors in our theoretical framework.

### Social Cues

Although many researchers have studied the effects of social cues (Tanis & Postmes, 2003; Walther et al., 2005; Wang et al., 2010), few have explicitly defined the term. Social cues can serve as affordances (Sundar et al., 2015), reflect physical attractiveness (Antheunis & Schouten, 2011), or be conceived as contextual information (Sproull & Kiesler, 1986). Here we specifically adopt Fiore et al.'s (2013) definition of cues as "biologically and physically determined features salient to observers because of their potential as channels of useful information" (p. 2). Fiore and colleagues distinguished social cues and social signals, where social cues can be understood as physical or behavioral features displayed by a social actor and social signals are the meaningful interpretations of these social cues by the perceivers. This is an important stepping-off point because social cues have been used interchangeably with social signals in some literature (Wiltshire et al., 2014). Examples of social cues include a social actor's voice, humanlike appearance, and eye gaze, whereas social signals include perceivers' translation of these and other cues into an understanding of the social actor's emotion (e.g., indicated by the actor's smiling face), attention (e.g., indicated by eye contact), empathy (e.g., indicated by hugs or language), and so on.

Based on this distinction, although the early CASA researchers manipulated computers to have personalities, genders, or conversation patterns, what they actually controlled were various sets of social cues. For instance, in their manipulation of gendered computers, Nass et al. (1997) assigned a female voice and a male voice to the computers and found that a female computer was perceived to be more familiar with love and relationships, whereas a male computer was more knowledgeable about technical subjects. Similarly, when operationalizing computer personalities, Nass and Lee (2001) controlled the levels of speech rate, volume, fundamental frequency, and pitch range to differentiate introverted and extroverted computers.

Although the role of social cues has been examined in much of the CASA literature, one question that Nass and Moon (2000) raised in their suggestions for future research is whether there exist some dimensions of a computer that are more likely to evoke social responses compared to others, and how different combinations of them would impose additive or synergetic influence on social responses. Although no sufficient findings have been established to advance a hierarchy of social cues which indicates the distinct power of each single cue over users' presence experiences and social responses, some scholarship at least organizes and compares various pairs of social cues. For example, Nass and Steuer (1993) examined the effects of different voices and of the same or different computer boxes producing the voices. They found that the voice manipulation had greater effects than the box one in predicting perceived accuracy and fairness of an evaluation session in human-computer interaction. Reeves and Nass (2000) also noted that sight and sound dominate human perception and play a more important role than other senses that can relay social cues like smell. Thus, it can be inferred that there should be a group of cues that are more likely to generate individuals' social perception than others. This postulation is consistent with the cognitive miser theory (Fiske & Taylor, 1991), based on which researchers suggest that how a person is construed is determined by the quality of available visual inputs. The cognitive process of person perception may occur automatically when critical cues such as facial features are available to the perceiver, but the same process may not be activated without the presence of these cues (Gauthier & Tarr, 1997; Martin & Macrae, 2007). Therefore, in the process of activating users' medium-as-social-actor presence and social responses, what matters is not only the quantity of social cues, but also the quality of social cues.

Based on prior research, we identify two groups of social cues, primary cues and secondary cues, to refer to their different effects on users' social reactions. Primary cues are those that are most salient and central to humans' perception of socialness. Responses to these primary cues are based on individuals' evolutionary bias toward humanlike (or animal-like) characteristics. Each primary social cue should be sufficient but not necessary to evoke social responses. By contrast, secondary cues are those that are less salient and less central to humans' perception of socialness. They are neither sufficient nor necessary to evoke social responses.

## Primary Cues

We include face, eye gaze, gesture, human-sounding voice, and humanlike or animal-like shape as major examples of primary cues. These cues have been found to evoke humans' evolutionarily based responses. For instance, researchers have found that infants are sensitive to fearful, surprising, and angry *facial expressions* (Kobiella, et al., 2008; Schmidt & Cohn, 2001; Serrano et al., 1992). Paredolia, which refers to the illusion in which people see faces in inanimate objects (e.g., taking a car's headlights as eyes and its bumper or grill as mouth) (Takahashi & Watanabe, 2013), also supports the idea that the face is a powerful cue in triggering social perception. Past research has corroborated that including a face as a social cue can lead participants to perceive computer agents as more trustworthy, persuasive, and positive (Gong, 2008; Shamekhi et al., 2018).

As an important social feature of a face, *eye gaze* is another evolutionarily significant characteristic of human beings. Eye gaze delivers the social signals of attention, emotion, or acknowledgment (Andrist et al., 2015; Fink & Penton-Voak, 2002). Past research has shown that even 12-month-old infants can follow the gaze direction of robots (Okumura et al., 2013).

Another social cue that serves as a basic element of our daily nonverbal interaction is *gestures* (Krauss et al., 1996). Johansson (1973) found that our visual perception of biological motion is keen enough to identify gestures that are made up of only 10 to 12 bright spots representing different human body joints. Prior literature has suggested that technologies designed with movable arms, hands, and/or bodies that imply intentions, motives, mental states, and social rituals can evoke our social perception (Salem et al., 2013). For instance, letting users shake hands with the robot *NAO* before negotiating with it led to more cooperation between users and robots (Bevan & Fraser, 2015).

The other two primary cues, *human-sounding voice* and *human- or animal-like shape*, have already received much attention in the field of HMC. Many studies have demonstrated that human-sounding voice, whether actual or synthetic, has greater effects than machine-sounding speech in evoking social responses (Chérif & Lemoine, 2019; Chiou et al., 2020; Xu, 2019). The human voice has especially been perceived as a natural and powerful cue and can more easily encourage individuals' learning performances compared to a machine voice (Mayer, 2014; Nass & Brave, 2005). Interacting with Siri is an example of how a technology paired with mere vocal cues can raise individuals' involuntary awareness of a social entity. Fortunati (2003) used the phrase "the body reaches where the voice does" to highlight the idea that voice serves as an extension of the body in our daily interaction (p. 62).

Designing humanlike or animal-like shape into technologies has been effective in forging an attractive appearance. For instance, Pixar's mascot Luxo is designed to have a head and a body to increase its popularity. Martini et al. (2016) found that the degree of robots' humanlike appearance has a positive relationship with individuals' attribution of intentionality to the robots. Hinds et al. (2004) found that participants felt less responsible for a task when collaborating with a humanlike robot than with a machinelike robot partner, meaning that participants attributed more trust to the humanlike one than the machinelike one. Using fMRI, research has further suggested that even observing animal-like shadows created by finger movements on a screen activates the same brain activity as verbal communication, implying that perceiving animal shadows involves the same physiological reaction as decoding human speech (Fadiga et al., 2006).

One might argue that humans shifting from acceptance to revulsion when a technology appears highly humanlike but falls short of a perfect human replica (Mori et al., 2012) counters the classification of humanlike shape as a primary cue. While it is true that users' affinity for the machine may plunge at the occurrence of this "uncanny valley," experiencing eeriness itself is a manifestation of strong medium-as-social-actor presence, as it shows that humans develop strong emotional responses to a machine that approximates human appearances. Mori et al. (2012) postulated that the eerie sensation could be due to the human instinct that protects us from potential dangers, including members of different species appearing to be humanlike, which supports the idea that our responses to humanlike technologies are sensitive and intuitive.

## Secondary Cues

Compared to primary cues that evoke evolutionary-based responses, secondary cues have less power in activating users' social perception and responses. Below we identify a few representative examples of secondary cues.

Whereas *human or animal size* could be a contributing factor to users' social responses to technologies (Duffy, 2003; Takayama & Pantofaru, 2009), it is not always evocative. Walters et al. (2009) found that while in single trials a short robot led participants to allow for a closer approach distance than a taller robot, the effect faded away in repeated trials over three weeks. Additionally, when explaining the uncanny valley effects, Mori et al. (2012) used the Bunraku puppet to indicate that audiences tend to ignore the size of the puppet but concentrate on its appearances.

As a basic human means of exchanging social information (Dunbar, 2004), *language use* has been found to trigger social perception and responses (Sah & Peng, 2015; Xu, 2020). However, it is classified as a secondary cue here because it does not always lead to strong social responses. For instance, a water bottle with the label "drink me" may induce greater levels of social reactions than one with the label "drink it." Thus, the effects depend on the variations in language. Informal language, warm (i.e., friendly and conversational) language, and language with vocal fillers and self-referential statements have been found to be more evocative than formal language, cold language, and language without paralinguistic cues (Goble & Edwards, 2018; Hoffmann et al., 2020; Sah & Peng, 2015).

*Motion* attracts our attention and may influence our social responses (Reeves & Nass, 2000). The Heider-Simmel (1944) experiment suggests that even simple dots that randomly move on a computer screen can be interpreted to have intentions. So can different forms of automatic doors that move with different speeds and trajectories (Ju & Takayama, 2009). However, movements have strong effects on social perception only when they become symbolic or conversational (Krauss et al., 1996). That is, movements that reveal communicators' purposes, mental states, and adherence to social rituals are more likely to evoke social reactions than random movements (Hoffman & Ju, 2012; Xu, 2019). For instance, Fiore et al. (2013) found that an *iRobot Ava* that gave way to participants as it traveled across a hallway evoked stronger social presence than one that did not yield to participants. Thus, depending on their speed, trajectory, frequency, and social meanings, a technology's motion cues may trigger different levels of social perception (Mori et al., 2012).

Compared to a human-sounding voice, a *machine-sounding voice* is generally perceived as unnatural and unpleasant (Gong & Lai, 2003). A machine-sounding voice has generally been found to be inferior to a human voice in evoking medium-as-social-actor presence experience (Chérif & Lemoine, 2019). Moreover, the level of social responses evoked by a machine's voice, or other symbolic sounds, depends on the relationship between the technology and its users. In our daily interaction with technologies, alarm system voices; smartphone ringtones; SMS, calendar, and other alerts and reminders; the myriad sounds built into Windows, iOS, and other computer operating systems; and even the ticking and chiming of mechanical clocks all suggest the potential interpretation of a social actor for users who have come to understand, be comfortable with, and rely on them.

## Social Signals

Beyond the primary and secondary cues, technologies can present social cues that can be interpreted as social signals in HMC (Fiore et al., 2013; Streater et al., 2012). That is, single or combinations of primary and/or secondary social cues can constitute abstract human characteristics including personalities, identities, and so on. Perceiving these social signals may require prolonged exposure to or interaction with the technologies. Contextual factors may also help individuals sense these abstract human characteristics. Below we provide some examples of social signals that are composed of social cues.

The ability of a media technology to be *responsive* or *interactive* should be sufficient on its own to trigger users' social perception and responses (Kim & Sundar, 2012). But presenting interactivity requires technologies to demonstrate a constellation of cues, which may include a physically embodied robot's eye contact, nodding, and smiling, or a chatbot's timing of response, message contingency, and so on (Jung et al., 2014; Lew et al., 2018).

*Perceived personality* and *social identity* of a media technology also requires a constellation of cues. A systematic review (Mou et al., 2020) suggests that personalities of machines can be operationalized via cues such as speech styles, vocal features, movements, and proximity. In addition, social identity can be easily perceived based on minimal, but meaningful, cues. For instance, Eyssel and Kuchenbrandt (2012) assigned different names to two robots, one German name and one Turkish name. Participants in the group with the same-nationality robot were more likely to evaluate the robot's performances as positive (Hogg & Abrams, 1988; Kuchenbrandt et al., 2013).

Some people (and animals) in our lives become our *companions*, and some media technologies (e.g., Tamagotchi, books, smartphones) can provide the same sense of social connection. Early uses and gratifications research suggested that people's ritualized use of television is related to higher affinity with the television itself rather than the content it presents (Rubin, 1983). The physical presence of a technology itself is an essential cue and the additions of several others (e.g., sound, touch) can enhance the perceived social companionship.

All human beings are unique; aside from genetic and physical differences we all have different life experiences (Orwig, 2014; Turkle, 2012). We associate *uniqueness* with other people and value it in our interpersonal relationships (Eastwick & Hunt, 2014). Many technologies can also be perceived as being unique, either because they come to us with these features or are designed or adapted for personalization. For example, many personal and mobile computers, virtual assistants, and other technologies ask users to provide a name for them; some people install custom ringtones and "skins" for electronic devices, and location-based mobile systems can provide customized context-aware information to assist users with decision-making (Choi et al., 2017). Even iRobot's Roomba vacuum cleaners can be personalized with a "skin toolkit" to increase users' acceptance of the device and their commitment to use it more (Sung et al., 2009).

A combination of social cues may cause individuals to infer the *health status* or the *life span* of technologies (Lechelt et al., 2020). When an old technology (e.g., mechanical clock, television, computer, smartphone, robot, etc.) begins to wear out or falter or is damaged in an accident, the owner may take it to a "doctor" (repair person) and wait with worry and impatience for a report on whether and how it will "recover." Even when technologies lose their functional value, or the developers stop supporting their use (e.g., the shutdown of servers that supported the Jibo Robot), the owners of the technologies may still find social value in keeping them as companions or for reimagined uses (see Lechelt et al., 2020).

Other cues and signals that could evoke medium-as-social-actor presence and social responses include olfactory cues (Chen, 2006), haptic cues (Blakemore, 2016; Li et al., 2017), interpersonal distance (Syrdal et al., 2006; Takayama & Pantofaru, 2009; Walters et al., 2009), and degree of flexibility (Duffy & Zawieska, 2012). It should be noted that even within a single cue, there exists significant components that evoke different effects. For example, vocal cues contain volume, pitch, tone, vocal outbursts, and so forth (Vinciarelli et al., 2009).

## Individual Differences and Contextual Factors

We have suggested that social cues differ in their power over users' social responses. These social cues can further be converted into social signals that act as perceived human characteristics such as personalities, identities, and so on. As part of the MASA paradigm, here we further suggest the importance of individual differences and contextual factors that likely play a role in users' medium-as-social-actor presence and social responses.

*Anthropocentrism* is "the tendency of individuals to perceive the world from a human-centered perspective, in which humankind is the most significant of all entities" (Nass et al., 1995, p. 229). A person high in anthropocentrism is less likely to believe that technologies can, and should, take on physical and psychological attributes of people and occupy human social roles, which would likely suppress their medium-as-social-actor presence.

The CASA research has suggested that individuals' own *personalities* play a role when interacting with technologies; for example, a robot that demonstrated a complementary personality (Lee et al., 2006). Related to users' personalities, individuals with greater desire for social interaction are more likely to modulate their presence experiences by looking for social cues (Gardner et al., 2005) and perceiving social actors when they use technologies (Epley et al., 2007).

Epley et al. (2007) argued that children are more likely than adults to ascribe human emotions to external entities (e.g., angry clouds, happy sun). Flavell et al. (1990) found that children are more likely to perceive mediated actors or objects as "real." The same logic applies to medium-as-social-actor presence. In her book *Alone Together,* Turkle (2012) observed children's interaction with the toy Furby and noticed that children believe that Furby can burp, understand language, and may one day "die." All these observations suggest that *age* is an influential factor in our responses to technologies (Edwards et al., 2019).

Beyond these factors, individuals differ in their *knowledge* about, *experience* with, and *ability to think critically* about technologies, which should lead to different levels of medium-as-social-actor presence. For example, Nass and Moon (2000) suggested there was no evidence that computer experts are immune to social responses to computers. Johnson and colleagues (2004) found that those with more computer use experiences reported more positive affect in response to a computer's flattery and more favorable judgments of the computer. In addition, Lee (2010) found that people who were more analytical and rational were less likely to respond to flattery effects of computers compared to those who were intuitive and experiential.

Prior literature has suggested the potential of many other individual differences in leveraging our medium-as-social-actor presence experiences. For example, Salem et al. (2013) found that participants preferred robots that occasionally performed incorrect gestures over those that always performed perfect ones, implying that our *tolerance of imperfection* may help determine our social attitudes toward technologies. For people who have unrealistically high *expectations* for the performance of technologies, every disappointment highlights the true nature of the technologies, so these people will be less likely to have strong social responses to them (Paepcke & Takayama, 2010; Waddell, 2018). Likewise, individuals' *willingness to suspend disbelief* may determine whether we treat a machine more as a tool or as a social entity (Duffy & Zawieska, 2012). Moreover, individuals with different *attachment styles* diverge in their intention to seek social cues and form social relationships with other social actors (Cole & Leets, 1999; Epley et al., 2007). Other relevant individual differences include gender, self-esteem, tolerance of uncertainty, tendency to make attribution errors, and so on (Epley et al., 2007; Nass et al., 1995; Rosen & Knäuper, 2009).

As with other types of presence (Slater & Wilbur, 1997; Won et al., 2015), a wide variety of factors related to the context of our exposure to and interactions with technologies may impact our experiences of medium-as-social-actor presence. These include, but are by no means limited to, the nature of the activity or task involved, the setting (e.g., public or private), the number of people present and our relationships with them, the amount and accuracy of information available to us regarding the nature of the technology, and even the time of day. More broadly, aspects of our culture are important factors (Bartneck et al., 2007). These include the prevailing attitudes toward technologies such as robots assuming social roles (Nass et al., 1995), the degree of industrialization that provides exposure to various technologies (Epley et al., 2007), the culture's tolerance of privacy invasion and data sharing (Nitto et al., 2017), and the subtle ways languages not only reflect but potentially guide our perceptions. To illustrate the power of language norms, note that in English at least, clocks have "faces" and "hands," chairs have "arms," needles have "eyes," streets have "shoulders," and computers and many other technologies can "die" (if we don't "kill" them when they don't "cooperate").

## MASA Mechanisms: Unifying Two Explanations

So far we have presented the roles of primary cues, secondary cues, and social signals in the MASA paradigm. And we've discussed how, along with these factors, individual differences and contextual elements may further moderate the breadth and depth of users' social responses. Below we discuss two major explanatory mechanisms that have been supported in prior literature on CASA: mindlessness and anthropomorphism. Instead of viewing the two mechanisms as competing explanations for users' social perceptions and responses, we suggest in the MASA paradigm that these two mechanisms can be unified and account for different social response scenarios.

Mindlessness refers to the explanation that people are naturally oriented to the social rather than the asocial cues of technologies (Langer, 2000; Nass & Moon, 2000). This may be based on evolutionarily based cognitive traits or learned from cues having been demonstrated repetitively in interpersonal communication, so that individuals "mindlessly (and) prematurely commit to overly simplistic scripts drawn in the past" (Nass & Moon, 2000, p. 83). In contrast, anthropomorphism refers to a more active phenomenon, "the tendency to imbue the real or imagined behavior of nonhuman agents (e.g., animals, nature, gods, and mechanical or electronic devices) with humanlike characteristics, motivations, intentions, or emotions" (Epley et al., 2007, p. 864).

While both mindlessness and anthropomorphism seem to be logical explanations for people's social perception and responses, both of them have been found to have limitations. Regarding mindlessness for instance, Fischer and colleagues (2011) noticed that some participants laughed when receiving a robot's greetings, indicating that they found something odd and amusing about the interaction with the robot. Mou and Xu (2017) found that participants demonstrated two different personalities in interactions with chatbots versus humans. These studies suggest that users' responses to machines are not as mindless and spontaneous as to humans. Anthropomorphism cannot account for all CASA findings either. For example, Lee (2010) found that compared to text-only conditions, computer interfaces with anthropomorphic cartoon characters enhanced the social attractiveness and the trustworthiness of the computer, but they did not amplify flattery effects, which challenges the prediction that more anthropomorphic characters would facilitate stronger social responses (Lee, 2010).

The two explanatory mechanisms are seemingly antagonistic, as mindlessness occurs "without extensive thought or deliberation" (Moon, 2000, p. 325), and anthropomorphism "involves the thoughtful, sincere belief that the object has human characteristics" (Nass & Moon, 2000, p. 93). However, a closer investigation into the role of technology cues may parse out the interrelationship between them. Specifically, Nass and Moon (2000) argued that to elicit mindlessness, an object must exhibit "enough cues" to bring forth social responses (p. 83), which implies that mindlessness may not occur when the cues are not evident or sufficient. On the other hand, anthropomorphism emphasizes the ascription of human mental or emotional states to nonhuman agents ranging from imagined ghosts to computer-generated dots (Duffy, 2003; Morewedge et al., 2007), which implies that even when objects are not designed with social cues (e.g., dots, clouds), anthropomorphism may occur as humans can mindfully assign human attributes to these nonhuman agents. Considering that Kim and Sundar (2012) noted that anthropomorphism can occur mindlessly or mindfully, we use mindless anthropomorphism and mindful anthropomorphism to unify the explanations for medium-as-social-actor presence and social responses. Specifically, mindless anthropomorphism should have more explanatory power when media technologies display cues that are of high quantity and high quality. In other words, if technologies demonstrate a group of primary social cues at the same time (e.g., Ishiguro's Gemonoid [Nishio et al., 2007] or Samsung's virtual human Neon [Vincent, 2020)]), users may not help having mindless, intuitive, and spontaneous responses to the technologies as these cues are natural, nuanced, and powerful. By contrast, when people experience strong medium-as-social-actor presence with technologies that are sufficient in neither the quality nor the quantity of cues, mindful anthropomorphism should be better at explaining the experience, as such scenarios indicate that individuals deliberately and thoughtfully attribute human characteristics to these technologies with limited cues (e.g., children imagining that their toys can speak to each other, drivers assigning names to their cherished cars) (see Figure 1).

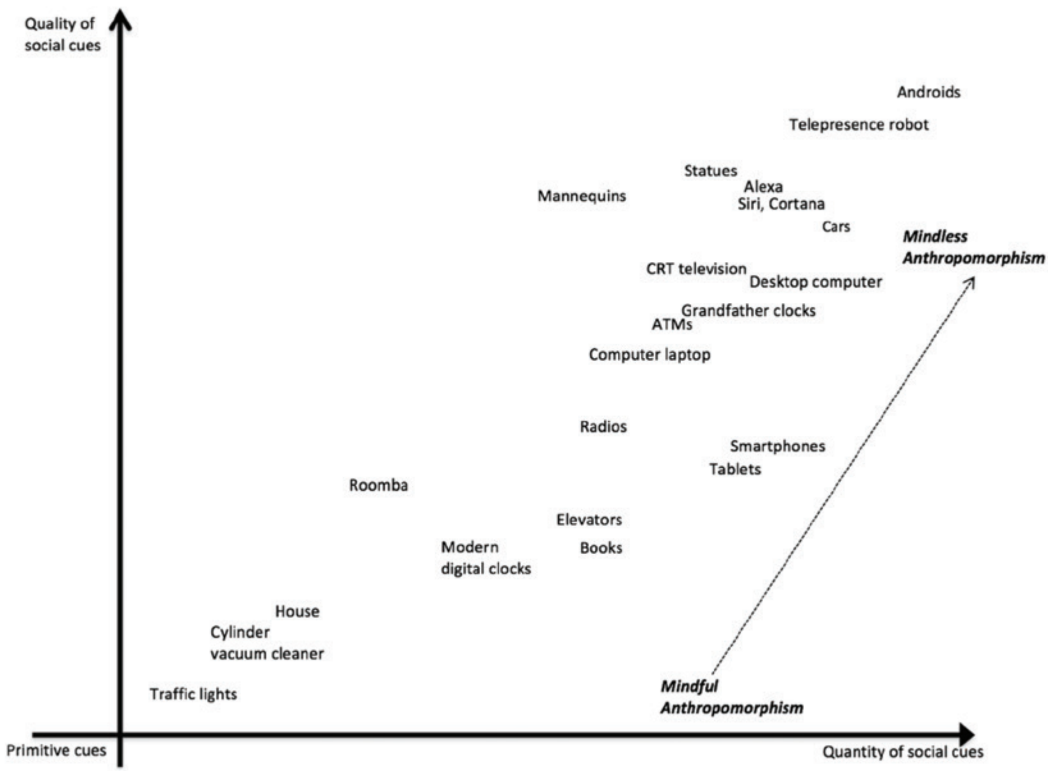



**FIGURE 1 The Relationships Among Social Cues, Technologies, and Explanatory Mechanisms**

Note: The placement of technologies in this figure represents only an approximation of the characteristics of prototypical examples of each technology.

The distinction between mindless and mindful anthropomorphism may be useful in understanding two related questions raised in CASA. The first question stems from the fact that although Nass and Moon (2000) endorsed the mindlessness explanation, CASA "fails to pinpoint precisely when and why mindless behavior will occur" (Nass & Moon, 2000, p. 96). Our mapping of mindless anthropomorphism and mindful anthropomorphism provides a reasonable, literature-based, and testable contention about which mechanism may be more explanatory depending on the nature and number of technology cues available to users.

The second question is related to the CASA assumption that technologies must demonstrate *enough* cues to elicit mindless social responses (Nass & Moon, 2000). What qualifies as enough, or in other words, how *social* the cues need to be to induce mindless responses, is unclear. Our categorization of cues as primary and secondary, along with the scope of application of mindless versus mindful anthropomorphism, provides a testable suggestion that in order to evoke mindless responses, the cues should ideally be sufficient in both quality and quantity. While research will need to establish the relative impacts of different combinations of high quality (primary) cues and secondary cues, the definition of primary cues suggests that quality is of paramount importance.

As with the effects of cues on medium-as-social-actor presence and social responses, the activation of mindless versus mindful processing is contingent upon a series of factors including the specific combination of cues, individual differences, and contextual factors. Thus, it is possible that some individuals may mindlessly respond to media with fewer social cues or mindfully respond to media that demonstrate more social cues. This is consistent with Fischer's (2011) speculation that individuals differ in their inclination to be mindless and Fussel and colleagues' (2008) finding that users demonstrate different degrees of anthropomorphism in their spontaneous responses to actual robots versus their carefully considered conceptions of robots.

We also readily acknowledge that other possible explanations for social responses that have received less attention in the CASA literature may become more useful in explaining our increasingly complicated and diverse media use practices. These include the source orientation explanation (Solomon & Wash, 2014), cognitive load explanation (Lee, 2008), and folk explanations of behavior (Malle, 1999).

## Evolutionary Foundation and Propositions of the MASA Paradigm

Above we have differentiated primary cues from secondary cues and parsed out the relationship between mindless and mindful anthropomorphism. Below we first suggest how the MASA paradigm and the social perceptions and responses it seeks to predict and explain are established upon three inter-related interpretations of socialness. Then we formally list the propositions of the MASA paradigm based on the arguments presented here and above.

In the CASA literature and so far in this paper the term social is treated as a "primitive" term (Chaffee, 2009) with a meaning that is widely understood; essentially something is social if it involves human perceptions and behaviors in response to other humans. But given its role in CASA and MASA, it is important to be clear about what is meant by this term. Here we provide three related conceptualizations of socialness. All three are consistent with the precepts of evolutionary psychology that suggest we have developed our cognitive, affective, and behavioral abilities via natural selection (Brewer, 2004; Sokol & Strout, 2007).

First, our brain is social. Theory of mind suggests that we can infer others' intentions, desires, and motivations based on their social cues (Adolphs, 2009; Banks, 2020; Frith & Frith, 2005). By predicting others' intention and actions, humans can adjust their own actions accordingly, and more effectively communicate and collaborate with others (Hare, 2007). Research has shown that in this inference process, the amygdala, an important structure in the medial temporal lobe of our brain, is activated to regulate our social perception and behavior (Adolphs, 2009). The tenet of theory of mind is consistent with the idea of egocentric modeling (Epley et al., 2007), which suggests that humans can only experience the world from their own perspective. Although we can speculate about others' intentions, we may make false predictions and cannot know exactly what it is like to be another animate or inanimate entity. Given the importance of interpreting social cues and the inherent challenges we face in doing so, the notion of the social brain means that no matter with what object we interact, we are at least somewhat oriented to perceive the object as social.

The second interpretation of being social indicates that our way of living is social. Here, social means being interdependent and cooperative (Brewer, 2004; Tomasello, 2014). Based on the perspective of adaptation, to resist environmental risks and avoid potential threats, primitives select to live in groups, make tools, and collaborate with each other. These means of coordination and collaboration have acted as a buffer between the individuals and actual and potential ecological changes (Caporael & Brewer, 1995). The instinct to avoid dangers is so imprinted in our adaptation that today we are still "hardwired" to monitor the world around us for threats (Shoemaker, 1996). It is consistent with Reeves and Nass's (1996) explanation that our brain has not evolved to distinguish mediated objects and nonmediated objects. It is also aligned with the findings regarding arousal and attention patterns in television viewers (Reeves & Thorson, 1986)—at some level viewers respond to movement and change even if the potential threat is only shown on an electronic display screen.

These two interpretations of humans' social orientation can help us understand how users can perceive technologies with even limited cues, including technologies not created or used for communication (i.e., not media technologies), as social. For example, imagine a person who sees a hammer for the first time and needs to figure out how to use it. Although a hammer serves primarily a simple, functional purpose, it can be perceived as social in that first, observing the shape of the hammer, feeling the weight of the head, and finding out that the handle is designed to be held by one hand, is a process of using minimal visual and haptic cues to predict the potential use of the technology. And second, using the hammer to strike other objects or pull out nails can be viewed as an interdependence between the user and the tool to accomplish a goal. Using technology is a form of collaboration that in our long history humans have enacted to attempt to improve our living conditions. So even though a technology may present limited cues and be designed for a simple purpose, it retains a subtle social element that we are hardwired to perceive.

The third interpretation of being social refers to our being adaptive and flexible. Research has found that although humans are not the only animals that infer each other's mental states, humans have evolved to develop the propensity for flexibility in using social cues during communication and cooperation with others, which is less evident for animals like chimpanzees (Hare, 2007). Demonstrating flexibility across different scenarios is likely to precipitate humans' cooperation process. This evolutionary characteristic is also inherent in our interaction with technologies. Some may argue if a study discovers that humans do not treat machines as humans, the study refutes the CASA proposition. However, the MASA paradigm holds that our social responses to technologies vary just as our social responses vary when we encounter and interact with children, foreigners, strangers, and friends (and among individuals within these categories) (Katagiri et al., 2001). As technologies transform to become multi-modal, ubiquitous, context-aware, and even invisible (Campbell, 2020; Fortunati, 1995; Ling, 2012), humans apply our ability to develop flexible and even unique social rules for interactions in all kinds of HMC scenarios. As an example, typing letters using a keyboard is a skill we learn so that we can interact with computers, but the behavior itself is ultimately a form of interdependence, which is considered the essence of being a social species (Caporael & Brewer, 1995; Tomasello, 2014).

We have explicated what *social* means in the MASA paradigm. We have also advocated for distinguishing the quantity of social cues and the quality of social cues presented by technologies. In addition, we explained how mindless and mindful responses can be dialectically applied to different technology use scenarios. Below we present a set of propositions based on the theoretical and empirical evidence discussed above. Specifically, Proposition 1 is related to the evolutionary foundation of the MASA paradigm. Propositions 2 to 6 are related to the different effects of social cues. Propositions 7 to 9 are based on the relationship between mindless and mindful anthropomorphism.

> **P1:** Every media technology has at least some potential to evoke medium-as-social-actor presence and corresponding social responses.

> **P2:** It is not only the social cues but also the combination of social cues, social signals, individual factors, and contextual factors that lead to medium-as-social-actor presence and corresponding social responses.

> **P3:** Some social cues are primary: Each is sufficient but not necessary to evoke medium-as-social-actor presence.

> **P4:** Some social cues are secondary: Each is neither sufficient nor necessary to evoke medium-as-social-actor presence.

> **P5a:** All other conditions being equal, individuals are more likely to experience medium-as-social-actor presence and socially respond to media technologies that display cues with more human characteristics (quality of cues).

> **P5b:** All other conditions being equal, individuals are more likely to experience medium-as-social-actor presence and socially respond to media technologies that display more social cues (quantity of cues).

**P6:** All other conditions being equal, the quality of cues (primary vs. secondary) has a greater role in evoking medium-as-social-actor presence and corresponding social responses than the quantity (number) of cues.

**P7:** Individuals vary in their tendency to perceive and respond to media technologies as social actors.

**P8**: Individuals' social responses to media technologies can occur with either mindless or mindful processing.

**P9a:** All other conditions being equal, media technologies that display more cues (quantity of cues) are more likely to lead individuals to *mindlessly* perceive them as social actors. Conversely, all other conditions being equal and *given the same level of social responses*, media technologies that display fewer cues (quantity of cues) are more likely to lead individuals to *mindfully* perceive them as social actors.

**P9b:** All other conditions being equal, media technologies that display cues with more human characteristics (quality of cues) are more likely to lead individuals to *mindlessly* perceive them as social actors. Conversely, all other conditions being equal and *given the same level of social responses*, media technologies that display cues with fewer human characteristics (quality of cues) are more likely to lead individuals to *mindfully* perceive them as social actors.

## Contributions, Scope, and Future Research

The MASA paradigm follows the major tenet of the CASA paradigm and expands it to more media technologies, including emerging and future technologies. The MASA paradigm can make the following contributions. First, the original CASA paradigm and the Media Equation described a series of experiments on users' social responses to computers and televisions. They did not explicitly list what propositions scholars could rely on to derive additional research questions. The MASA paradigm lists nine propositions here so that future research can test, refine, and enhance the theoretical framework in its predictive power, explanatory power, falsifiability, heuristic value, and so on (Chaffee & Berger, 1987).

Second, most prior research has primarily focused on the effects of single social cues or the cumulative effects of social cues, while the quality of social cues—whether they are central to our perception of socialness—has not been systematically examined. The MASA paradigm suggests that there exists a hierarchy of social cues that represents their potential to evoke social perceptions and responses. With knowledge about the quality of social cues, designers and developers may scrutinize different effects of social cues and find the optimal combination of cues to create positive use experiences.

Third, the paradigm identifies the roles of individual differences and contextual factors in generating users' medium-as-social-actor presence and social responses. It provides an analysis of the whole communication scenario in which individuals are perceivers, technologies are manifestations of social cues, and contexts guide perceptions and responses, with all of these elements influential in evoking medium-as-social-actor presence and social responses (see Figure 2). Future research can test the moderating roles of particular individual differences and contextual factors while investigating users' social reaction to different sets of technology cues.

Fourth, the paradigm has articulated an evolutionary base and explained what it means to have "social" perceptions and responses. Social responses do not necessarily equate to human responses. Social perceptions and social responses occur because we have a social brain, live a social life, and are used to making adaptations to be cooperative and interdependent. Based on this theoretical foundation, we have attempted to provide answers to Nass and Moon's (2000) questions about (1) what dimensions of technologies are more powerful in evoking social responses, (2) when mindless processing is more likely to occur, and (3) how "social" a technology should be to bring forth mindless social responses. We believe that the current MASA paradigm can be useful in future research on social responses to technologies and we encourage theoretical refinements in the paradigm that can bring greater understanding of human behaviors with respect to technology and make possible the practical benefits of improved product design and user experience.

Despite these contributions, some may wonder about the technologies to which the MASA paradigm applies and whether the term MASA could be replaced by TASA (i.e., Technologies are Social Actors). Here we refer to a technology as a "machine, device, or other application of human industrial arts" (ISPR, 2000, presence defined). We refer to media technologies as "artifacts or devices used to communicate or convey information" (Lievrouw & Livingstone, 2006, p. 23). In explaining that the MASA paradigm is rooted in evolutionary psychology we noted that humans are at least somewhat oriented to perceive technologies as social. Therefore, as part of the MASA paradigm, we acknowledge that people's social responses may occur to all technologies (sometimes even beyond technologies), especially when people mindfully and consciously anthropomorphize those technologies that are not designed to be used for social purposes. As media and communication scholars, however, we concentrate more on media technologies, including both traditional and emerging ones such as books, televisions, computers, smartphones, tablets, and so on. More importantly, we call it the MASA paradigm because this term reflects a shift of focus

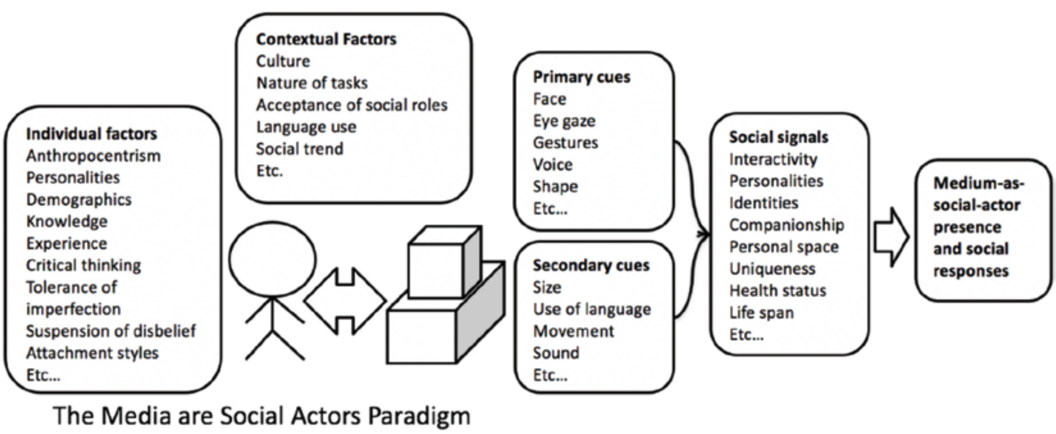



**FIGURE 2 The Media Are Social Actors paradigm**

from media as channels for message transmission to media as interlocutors or communicators (Gunkel, 2012). Although the role of media technologies may neither be entirely of a communication channel nor of a communication terminal in the current media landscape (Guzman, 2018), the MASA paradigm is targeted at the phenomena in which users perceive media technologies to be independent social entities, and partially or sometimes fully overlook their nature as mere machines or communication channels.

Several paths are available for researchers to investigate the validity of the MASA paradigm and refine and expand it. First, researchers should conduct empirical studies to verify the nine preliminary propositions. Studies can explore whether there exists a hierarchy of social cues that reliably elicit different degrees and/or particular types of social responses. Although few methods can provide direct evidence to investigate propositions that require differentiating users' mindless and mindful processing, researchers can use a combination of methods and measures, including both objective and subjective ones (e.g., fMRI and EEG versus interviews) and real-time and retrospective ones (e.g., secondary task reaction time and think aloud protocols versus post-exposure memory tests and questionnaires), to look for the convergence of results.

Second, researchers should look beyond CASA studies to test whether and how findings and predictions from other interpersonal communication theories and computer-mediated communication (CMC) theories can be adapted to the MASA paradigm. Some interpersonal communication theories researchers might examine include Communication Accommodation Theory (Giles, 2008), Communication Privacy Management Theory (Petronio & Durham, 2008), Expectancy Violation Theory (Burgoon & Hale, 1988), Uncertainty Management Theory (Bradac, 2001), Interpersonal Deception Theory (Burgoon et al., 2008), and the Theory of Imagined Interaction (Honeycutt, 2002). Some CMC theories or models that have been applied but could be explored further include the Social Identity model of De-individuation Effects (SIDE) (Reicher et al., 1995) and Social Information Processing Theory (Walther, 1996).

Third, researchers can further examine and add details to the psychological processes described in the MASA paradigm. For example, how do attention and memory influence medium-as-social-actor presence? What determines whether presence will lead to behavioral social responses? As researchers have not achieved consensus on the explanations for the phenomena captured in the CASA paradigm, more research should be conducted to account for the mental processes that underlie people's social perceptions of and responses to today's and tomorrow's diverse and evolving technologies.

Finally, researchers should explore and discuss with designers how they can apply the MASA paradigm and findings. Though most of the time richer media presentations can bring about better experiences, higher medium-as-social-actor presence (as with any type of presence) does not necessarily produce higher efficacy or popularity. Thus, designers should be fully aware of the advantages and disadvantages of their choices in applying the paradigm. They should also consider the larger ethical issues involved in creating technologies that lead users to perceive and treat media as social actors. As it is possible to view users of these technologies as victims of deception, unconscious responses, and the manipulation of presence (Biocca et al., 2003; Lombard, 2009), designing and marketing technologies that evoke medium-as-social-actor-presence and social responses should be based on thoughtful, informed, and ethical analyses.

## Author Biographies

**Matthew Lombard** (PhD, Stanford University) is co-founder and president of the International Society for Presence Research (ISPR, http://ispr.info) and an associate professor in the Department of Media Studies and Production at Temple University in Philadelphia, Pennsylvania, USA. His research centers on individuals' psychological and physiological processing of media presentations and experiences, with particular focus on the concept of (tele)presence. His work has appeared in academic journals including *Behaviour & Information Technology, CyberPsychology and Behavior, Human Communication Research, Human Technology, Journal of Computer-Mediated Communication, Journal of Communication,* and *Presence: Teleoperators and Virtual Environments.* For more detailed information please visit http://matthewlombard.com.

**Kun Xu** (PhD, Temple University) is an assistant professor of emerging media at University of Florida. His work focuses on the concepts of social cues and social presence in human-computer interaction, human-robot interaction, and computer-mediated communication.

https://orcid.org/0000-0001-9044-821X

## References

Abubshait, A., & Wiese, E. (2017). You look human, but act like a machine: Agent appearance and behavior modulate different aspects of human-robot interaction. *Frontiers in Psychology, 8*, 1393. https://doi.org/10.3389/fpsyg.2017.01393

Adolphs, R. (2009). The social brain: Neural basis of social knowledge. *Annual Review of Psychology*, *60*, 693–716.

Andrist, S., Mutlu, B., & Tapus, A. (2015). Look like me: Matching robot personality via gaze to increase motivation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 3603–3612). https://doi.org/10.1145/2702123.2702592

Antheunis, M. L., & Schouten, A. P. (2011). The effects of other-generated and system-generated cues on adolescents' perceived attractiveness on social network sites. *Journal of Computer-Mediated Communication*, *16*(3), 391–406. https://doi.org/10.1111/j.1083-6101.2011.01545.x

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior*, *85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Banks, J. (2020). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics, 12*(2), 403–414. https://doi.org/10.1007/s12369-019-00588-x

Bartneck, C., Reichenbach, J., & Carpenter, J. (2008). The carrot and the stick: The role of praise and punishment in human-robot interaction. *Interaction Studies, 9*(2), 179–203. https://doi.org/10.1075/is.9.2.03bar

Bartneck, C., Suzuki, T., Kanda, T., & Nomura, T. (2007). The influence of people's culture and prior experiences with Aibo on their attitude towards robots. *AI and Society, 21*(1), 217–230. https://doi.org/10.1007/s00146-006-0052-7

Bevan, C., & Fraser, D. S. (2015). Shaking hands and cooperation in tele-present human-robot negotiation. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 247–254). IEEE.

Biocca, F., Harms, C., & Burgoon, J. K. (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & virtual environments, 12*(5), 456–480. https://doi.org/10.1162/105474603322761270

Blakemore, E. (2016). Touching robots can turn humans on, study finds. *Washington Post.* https://www.washingtonpost.com/news/speaking-of-science/wp/2016/04/05/touching-robots-can-turn-humans-on-study-finds/

Bradac, J. J. (2001). Theory comparison: Uncertainty reduction, problematic integration, uncertainty management, and other curious constructs. *Journal of Communication, 51*(3), 456–476. https://doi.org/10.1111/j.1460-2466.2001.tb02891.x

Brewer, M. B. (2004). Taking the social origins of human nature seriously: Toward a more imperialist social psychology. *Personality and Social Psychology Review, 8*(2), 107–113. https://doi.org/10.1207/s15327957pspr0802_3

Burgoon, J. K., Blair, J. P., & Strom, R. E. (2008). Cognitive biases and nonverbal cue availability in detecting deception. *Human Communication Research, 34*(4), 572–599. https://doi.org/10.1111/j.1468-2958.2008.00333.x

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Cederberg, C., Lundeberg, M., & Allspach, L. (2000). Interactivity in human-computer interaction: A study of credibility, understanding, and influence. *Computers in Human Behavior, 16*(6), 553–574.

Burgoon, J. K., & Hale, J. L. (1988). Nonverbal expectancy violations: Model elaboration and application to immediacy behaviors. *Communication Monographs, 55*(1), 58–79. https://doi.org/10.1080/03637758809376158

Campbell, S. W. (2020). Cutting the cord: Social and scholarly revolutions as CMC goes mobile. *Journal of Computer-Mediated Communication, 25*(1), 101–110. https://doi.org/10.1093/jcmc/zmz021

Caporael, L. R., & Brewer, M. B. (1995). Hierarchical evolutionary theory: There is an alternative, and it's not creationism. *Psychological Inquiry, 6*(1), 31–34. https://doi.org/10.1207/s15327965pli0601_2

Carolus, A., Muench, R., Schmidt, C., & Schneider, F. (2019). Impertinent mobiles—Effects of politeness and impoliteness in human-smartphone interaction. *Computers in Human Behavior, 93*, 290–300. https://doi.org/10.1016/j.chb.2018.12.030

Chaffee, S. H. (2009). Thinking about theory. In D. W. Stacks & M. B. Salwen (Eds.), *An integrated approach to communication theory and research* (pp. 27–43). Lawrence Erlbaum.

Chaffee, S. H., & Berger, C. R. (1987). What communication scientists do. In C. R. Berger & S. H. Chaffee (Eds.), *Handbook of communication science* (pp. 99–122). Sage.

Chérif, E., & Lemoine, J. F. (2019). Anthropomorphic virtual assistants and the reactions of Internet users: An experiment on the assistant's voice. *Recherche et Applications en Marketing*, *34*(1), 28–47. https://doi.org/10.1177/2051570719829432

Chen, Y. (2006). Olfactory display: Development and application in virtual reality therapy. In *The 16th International Conference on Artificial Reality and Telexistence* (pp. 580–584).

Chiou, E. K., Schroeder, N. L., & Craig, S. D. (2020). How we trust, perceive, and learn from virtual humans: The influence of voice quality. *Computers & Education*, *146*, 103756. https://doi.org/10.1016/j.compedu.2019.103756

Choi, B., Kwon, O., & Shin, B. (2017). Location-based system: Comparative effects of personalization vs ease of use. *Telematics and Informatics, 34*(1), 91–102. https://doi.org/10.1016/j.tele.2016.04.011

Cole, T., & Leets, L. (1999). Attachment styles and intimate television viewing: Insecurely forming relationships in a parasocial way. *Journal of Social and Personal Relationships, 16*(4), 495–511. https://doi.org/10.1177/0265407599164005

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems, 42*, 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

Duffy, B. R., & Zawieska, K. (2012). Suspension of disbelief in social robotics. In *The 21st IEEE International Symposium on Robot and Human Interactive Communication* (pp. 484–489). IEEE.

Dunbar, R. I. M. (2004). Gossip in evolutionary perspective. *Review of General Psychology, 8*, 100–110. https://doi.org/10.1037/1089-2680.8.2.100

Eastwick, P. W., & Hunt, L. L. (2014). Relational mate value: Consensus and uniqueness in romantic evaluations. *Journal of Personality and Social Psychology, 106*(5), 728–751. https://doi.org/10.1037/a0035884

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social identity theory in human-robot interactions. *Computers in Human Behavior, 90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology, 51*(4), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

Fadiga, L., Craighero, L., Destro, M. F., Finos, L., Cotillon-Williams, N., Smith, A. T., & Castiello, U. (2006). Language in shadow. *Social Neuroscience, 1*(2), 77–89. https://doi.org/10.1080/17470910600976430

Fink, B., & Penton-Voak, I. (2002). Evolutionary psychology of facial attractiveness. *Current Directions in Psychological Science, 11*(5), 154–158. https://doi.org/10.1111/1467-8721.00190

Fiore, S. M., Wiltshire, T. J., Lobato, E. J. C., Jentsch, F. G., Huang, W. H., & Axelrod, B. (2013). Toward understanding social cues and signals in human-robot interaction: Effects of robot gaze and proxemic behavior. *Frontiers in Psychology, 4*, 1–15. https://doi.org/10.3389/fpsyg.2013.00859

Fischer, K. (2011). Interpersonal variation in understanding robots as social actors. In *The 2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 53–60).

Fischer, K., Foth, K., Rohlfing, K., & Wrede, B. (2011). Mindful tutors: Linguistic choice and action demonstration in speech to infants and a simulated robot. *Interaction Studies-Social Behaviour and Communication in Biological and Artificial Systems, 12*(1), 134–161. https://doi.org/10.1075/is.12.1.06fis

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill.

Flavell, J. H., Flavell, E. R., Green, F. L., & Korfmacher, J. E. (1990). Do young children think of television images as pictures or real objects? *Journal of Broadcasting & Electronic Media, 34*(4), 399–419. https://doi.org/10.1080/08838159009386752

Fortunati, L. (1995). *Gli italiani al telefono [Italians at the phone]*. Angeli.

Fortunati, L. (2003). Real people, artificial bodies. In L. Fortunati, J. E. Katz, & R. Riccini (Eds.), *Mediating the human body: Technology, communication, and fashion* (pp. 61–74). Lawrence Erlbaum Associates, Inc.

Frith, C., & Frith, U. (2005). Theory of mind. *Current Biology, 15*, 644–645.

Fussell, S. R., Kiesler, S., Setlock, L. D., & Yew, V. (2008). How people anthropomorphize robots. In *3rd ACM/IEEE International Conference on Human-Robot Interaction* (pp. 145–152).

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71–86. https://doi.org/10.30658/hmc.1.5

Gardner, W. L., Pickett, C. L., Jefferis, V., & Knowles, M. (2005). On the outside looking in: Loneliness and social monitoring. *Personality and Social Psychology Bulletin, 31*(11), 1549–1560. https://doi.org/10.1177/0146167205277208

Gauthier, I., & Tarr, M. J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research*, *37*(12), 1673–1682. https://doi.org/10.1016/S0042-6989(96)00286-6

Ghazali, A. S., Ham, J., Barakova, E., & Markopoulos, P. (2018). The influence of social cues in persuasive social robots on psychological reactance and compliance. *Computers in Human Behavior, 87*, 58–65. https://doi.org/10.1016/j.chb.2018.05.016

Giles, H. (2008). *Communication accommodation theory*. Sage Publications.

Goble, H., & Edwards, C. (2018). A robot that communicates with vocal fillers has… Uhhh… greater social presence. *Communication Research Reports, 35*(3), 256–260. https://doi.org/10.1080/08824096.2018.1447454

Gong, L. (2008). How social is social responses to computers? The function of the degree of anthropomorphism in computer representations. *Computers in Human Behavior, 24*(4), 1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Gong, L., & Lai, J. (2003). To mix or not to mix synthetic speech and human speech? Contrasting impact on judge-rated task performance versus self-rated performance and attitudinal responses. *International Journal of Speech Technology, 6*(2), 123–131. https://doi.org/10.1023/A:1022382413579

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication+ 1*, *1*(1), 1–25. https://doi.org/10.7275/R5QJ7F7R

Guzman, A. L. (Ed.). (2018). *Human-machine communication: Rethinking communication, technology, and ourselves*. Peter Lang Publishing, Incorporated.

Hare, B. (2007). From nonhuman to human mind: What changed and why? *Current Directions in Psychological Science*, *16*(2), 60–64. https//doi.org/10.1111/j.1467-8721.2007.00476.x

Heider, F., & Simmel, M. (1944). An experimental study of apparent behavior. *The American Journal of Psychology, 57*(2), 243–259. https://doi.org/10.2307/1416950

Hinds, P. J., Roberts, T. L., & Jones, H. (2004). Whose job is it anyway? A study of human-robot interaction in a collaborative task. *Human–Computer Interaction, 19*(1), 151–181. https://doi.org/10.1207/s15327051hci1901&2_7

Hoffman, G., & Ju, W. (2012). Designing robots with movement in mind. *Journal of Human Robot Interaction*, *1*(1), 78–95. https://doi.org/10.5898/JHRI.3.1.Hoffman

Hoffmann, L., Derksen, M., & Kopp, S. (2020, March). What a pity, Pepper! How warmth in robots' language impacts reactions to errors during a collaborative task. In *2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 245–247). https://doi.org/10.1145/3371382.3378242

Hogg, M. A., & Abrams, D. (1988). *Social identifications: A social psychology of intergroup relations and group processes*. Routledge.

Honeycutt, J. M. (2002). *Imagined interactions: Daydreaming about communication*. Hampton Pr.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction. Observations on intimacy at a distance. *Psychiatry, 19,* 215–229. https://doi.org/10.1080/00332747.1956.11023049

International Society for Presence Research (ISPR). (2000). *The concept of presence: Explication statement.* ISPR. https://ispr.info/about-presence-2/about-presence/

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics, 14*(2), 201–211. https://doi.org/10.3758/BF03212378

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

Ju, W., & Takayama, L. (2009). Approachability: How people interpret automatic door movement as gesture. *International Journal of Design, 3*(2), 1–10.

Jung, M. F., Martelaro, N., Hoster, H., & Nass, C. (2014). Participatory materials: Having a reflective conversation with an artifact in the making. In *Proceedings of the 2014 Conference on Designing Interactive Systems* (pp. 25–34). https://doi.org/10.1145/2598510.2598591

Kanda, T., Shiomi, M., Miyashita, Z., Ishiguro, H., & Hagita, N. (2009, 2008). An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction* (pp. 173–180). https://doi.org/10.1145/1514095.1514127

Katagiri, Y., Nass, C., & Takeuchi, Y. (2001). Cross-cultural studies of the computers are social actors paradigm: The case of reciprocity. In M. Smith, G. Salvendy, & D. Harris (Eds.), *Usability evaluation and interface design: Cognitive engineering, intelligent agents, and virtual reality* (pp. 1558–1562). Lawrence Erlbaum.

Kim, Y., & Sundar, S. S. (2012). Anthropomorphism of computers: Is it mindful or mindless? *Computers in Human Behavior, 28*(1), 241–250. https://doi.org/10.1016/j.chb.2011.09.006

Kobiella, A., Grossmann, T., Reid, V. M., & Striano, T. (2008). The discrimination of angry and fearful facial expressions in 7-month-old infants: An event-related potential study. *Cognition and Emotion, 22*(1), 134–146. https://doi.org/10.1080/02699930701394256

Krauss, R. M., Chen, Y., & Chawla, P. (1996). Nonverbal behavior and nonverbal communication: What do conversational hand gestures tell us? *Advances in Experimental Social Psychology, 28*(C), 389–450. https://doi.org/10.1016/S0065-2601(08)60241-5

Kuchenbrandt, D., Eyssel, F., Bobinger, S., & Neufeld, M. (2013). When a robot's group membership matters. *International Journal of Social Robotics, 5*(3), 409–417. https://doi.org/10.1007/s12369-013-0197-8

Langer, E. J. (2000). Mindful learning. *Current Directions in Psychological Science, 9*(6), 220–223. https://doi.org/10.1111/1467-8721.00099

Lechelt, S., Gorkovenko, K., Soares, L. L., Speed, C., Thorp, J. K., & Stead, M. (2020). Designing for the end of life of IoT objects. In *Companion Publication of the 2020 ACM Designing Interactive Systems* (pp. 417–420).

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, K. M., Peng, W., Jin, S. A., & Yan, C. (2006). Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human-robot interaction. *Journal of Communication, 56*(4), 754–772. https://doi.org/10.1111/j.1460-2466.2006.00318.x

Lew, Z., Walther, J. B., Pang, A., & Shin, W. (2018). Interactivity in online chat: Conversational contingency and response latency in computer-mediated communication. *Journal of Computer-Mediated Communication*, *23*(4), 201–221. https://doi.org/10.1093/jcmc/zmy009

Li, J. J., Ju, W., & Reeves, B. (2017). Touching a mechanical body: Tactile contact with body parts of a humanoid robot is physiologically arousing. *Journal of Human-Robot Interaction, 6*(3), 118–130. https://doi.org/10.5898/jhri.6.3.li

Lievrouw, L. A., & Livingstone, S. (2006). *Handbook of new media: Social shaping and consequences of ICTs.* Sage.

Ling, R. (2012). *Taken for grantedness: The embedding of mobile communication into society*. MIT Press.

Lombard, M. (2009). The promise and peril of telepresence. In C. Bracken & P. Skalski (Eds.), *Immersed in media: Telepresence in everyday life* (pp. 197–228). Routledge.

Lombard, M., & Ditton, T. (1997). At the heart of it all: The concept of presence. *Journal of Computer-Mediated Communication, 3*(2), JCMC321. https://doi.org/10.1111/j.1083-6101.1997.tb00072.x

Lombard, M., & Jones, M. T. (2015). Defining presence. In M. Lombard, F. Biocca, W. A. Ijsselsteijn, J. Freeman, & R. Schaevitz (Eds.), *Immersed in media: Telepresence theory, measurement, and technology.* Springer.

Malle, B. F. (1999). How people explain behavior: A new theoretical framework. *Personality and Social Psychology Review*, *3*(1), 23–48. https://doi.org/10.1207/s15327957pspr0301_2

Martin, D., & Macrae, C. N. (2007). A face with a cue: Exploring the inevitability of person categorization. *European Journal of Social Psychology, 37*(5), 806–816. https://doi.org/10.1002/ejsp.445

Martini, M. C., Gonzalez, C. A., & Wiese, E. (2016). Seeing minds in others–Can agents with robotic appearance have human-like preferences? *PloS One, 11*(1), e0146310. https://doi.org/10.1371/journal.pone.0146310

Mayer, R. E. (2014). *The Cambridge handbook of multimedia learning*. Cambridge University Press.

Moon, Y. (2000). Intimate exchanges: Using computers to elicit self-disclosure from consumers. *Journal of Consumer Research, 26*(4), 323–339. https://doi.org/10.1086/209566

Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology*, *93*(1), 1–11.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100.

Mou, Y., Shi, C., Shen, T., & Xu, K. (2020). A systematic review of the personality of robot: Mapping its conceptualization, operationalization, contextualization and effects. *International Journal of Human-Computer Interaction, 36*(6), 591–605. https://doi.org/10.1080/10447318.2019.1663008

Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior, 72,* 432–440. https://doi.org/10.1016/j.chb.2017.02.067

Nass, C. (2004). Etiquette equality: Exhibitions and expectations of computer politeness. *Communications of the ACM, 47*, 35–37. https://doi.org/10.1145/975817.975841

Nass, C., & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. MIT Press.

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human Computer Studies, 45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied, 7*(3), 171–181. https://doi.org/10.1037/1076-898X.7.3.171

Nass, C., Lombard, M., Henriken, L. & Steuer, J. (1995). Anthropocentrism and computers. *Behaviour and Information Technology, 14*(4), 229–238. https://doi.org/10.1080/01449299508914636

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

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., & Steuer, J. (1993). Voices, boxes, and sources of messages: Computers and social actors. *Human Communication Research, 19*(4), 504–527. https://doi.org/10.1111/j.1468-2958.1993.tb00311.x

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computer are social actors. *Human Factors in Computing Systems, 94*, 72–78. https://doi.org/10.1145/259963.260288

Nishio, S., Ishiguro, H., & Hagita, N. (2007). Geminoid: Teleoperated android of an existing person. In A. C. Philho (Ed.), *Humanoid robots: New developments* (pp. 343–352). InTech.

Nitto, H., Taniyama, D., & Inagaki, H. (2017). Social acceptance and impact of robots and artificial intelligence. *Nomura Research Institute, 211*, 1–15.

Nomura, T., & Kanda, T. (2015). Influences of evaluation and gaze from a robot and humans' fear of negative evaluation on their preferences of the robot. *International Journal of Social Robotics*, *7*(2), 155–164. https://doi.org/10.1007/s12369-014-0270-y

Okumura, Y., Kanakogi, Y., Kanda, T., Ishiguro, H., & Itakura, S. (2013). Can infants use robot gaze for object learning? The effects of verbalization. *Interaction Studies, 14*(3), 351–365. https://doi.org/10.1075/is.14.3.03oku

Orwig, J. (2014, September 24). More than 7 billion people populate earth and each one has a unique face – Here is why. *Business Insider*. http://www.businessinsider.com/why-are-human-faces-so-complex-and-unique-2014-9

Paepcke, S., & Takayama, L. (2010, March 2–5). Judging a bot by its cover: An experiment on expectation setting for personal robots. In *The 2010 5th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 45–52).

Perry, T. (2014). *Suitable technologies opens store staffed only by robots.* ISPR. http://ispr.info/2014/12/23/telepresence-robot-store-is-staffed-only-by-robots/

Petronio, S., & Durham, W. T. (2008). Understanding and applying communication privacy management theory. In D. Braithwaite & P. Schrodt (Eds.), *Engaging theories in interpersonal communication: Multiple perspectives* (pp. 309–322). Sage Publications.

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

Reeves, B., & Nass, C. (2000). Perceptual bandwith. *Communications of the ACM, 43*(3), 65–70. https://doi.org/10.1145/330534.330542

Reeves, B., & Thorson, E. (1986). Watching television: Experiments on the viewing process. *Communication Research, 13*(3), 343–361. https://doi.org/10.1177/009365086013003004

Reicher, S. D., Spears, R., & Postmes, T. (1995). A social identity model of deindividuation phenomena. *European Review of Social Psychology*, *6*(1), 161–198. https://doi.org/10.1080/14792779443000049

Rosen, N. O., & Knäuper, B. (2009). A little uncertainty goes a long way: State and trait differences in uncertainty interact to increase information seeking but also increase worry. *Health Communication, 24*(3), 228–238. https://doi.org/10.1080/10410230902804125

Rubin, A. M. (1983). Television uses and gratifications: The interactions of viewing patterns and motivations. *Journal of Broadcasting, 27*(1), 37–51. https://doi.org/10.1080/08838158309386471

Sah, Y. J., & Peng, W. (2015). Effects of visual and linguistic anthropomorphic cues on social perception, self-awareness, and information disclosure in a health website. *Computers in Human Behavior, 45*, 392–401. https://doi.org/10.1016/j.chb.2014.12.055

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human(-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics, 5*(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Schmidt, K. L., & Cohn, J. F. (2001). Human facial expressions as adaptations: Evolutionary questions in facial expression research. *Yearbook of Physical Anthropology, 44*, 3–24. https://doi.org/10.1002/ajpa.20001

Serrano, J. M., Iglesias, J., & Loeches, A. (1992). Visual discrimination and recognition of facial expressions of anger, fear, and surprise in 4- to 6-month-old infants. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology*, *25*(6), 411–425. https://doi.org/10.1002/dev.420250603

Shamekhi, A., Liao, Q. V., Wang, D., Bellamy, R. K., & Erickson, T. (2018, April). Face value? Exploring the effects of embodiment for a group facilitation agent. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (pp. 1–13). https://doi.org/10.1145/3173574.3173965

Shoemaker, P. J. (1996). Hardwired for news: Using biological and cultural evolution to explain the surveillance function. *Journal of Communication, 46*(3), 32–47. https://doi.org/10.1111/j.1460-2466.1996.tb01487.x

Short, J., Williams, E., & Christie, B. (1976). Theoretical approaches to differences between media. In *Social Psychology of Telecommunications* (pp. 61–66). Wiley.

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence: Teleoperators and Virtual Environments, 6*, 603–616. https://doi.org/10.1162/pres.1997.6.6.603

Sokol, R. I., & Strout, S. L. (2007). Understanding human psychology: The integration of social, evolutionary, and cultural studies. *Journal of Social, Evolutionary, and Cultural Psychology*, *1*(1), 1. https://doi.org/10.1037/h0099089

Solomon, J., & Wash, R. (2014). Human-what interaction? Understanding user source orientation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 422–426. https://doi.org/10.1177/1541931214581088

Sproull, L., & Kiesler, S. (1986). Reducing social context cues: Electronic mail in organizational communication. *Management Science, 32*(11), 1492–1512. https://doi.org/10.1287/mnsc.32.11.1492

Streater, J., Bockelman Morrow, P., & Fiore, S. (2012). Making things that understand people: The beginnings of an interdisciplinary approach for engineering computational social intelligence. In *The 56th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 22–26).

Sundar, S. S., Jia, H., Waddell, 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). John Wiley & Sons, Inc.

Sung, J., Grinter, R. E., & Christensen, H. I. (2009). "Pimp my Roomba": Designing for personalization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 193–196*)*. https://doi.org/10.1145/1518701.1518732

Syrdal, D. S., Dautenhahn, K., Woods, S., Walters, M. L., & Koay, K. L. (2006, Sept. 6–8). 'Doing the right thing wrong' - Personality and tolerance to uncomfortable robot approaches. In *15th IEEE International Symposium on Robot and Human Interactive Communication* (pp. 183–188). IEEE.

Takahashi, K., & Watanabe, K. (2013). Gaze cueing by pareidolia faces. *i-Perception, 4*(8), 490–492. https://doi.org/10.1068/i0617sas

Takayama, L., & Pantofaru, C. (2009, October). Influences on proxemic behaviors in human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems* (pp. 5495–5502). IEEE.

Tanis, M., & Postmes, T. (2003). Social cues and impression formation in CMC. *Journal of Communication, 53*(4), 676–693. https://doi.org/10.1111/j.1460-2466.2003.tb02917.x

Terzioğlu, Y., Mutlu, B., & Şahin, E. (2020, March). Designing social cues for collaborative robots: The role of gaze and breathing in human-robot collaboration. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 343–357). https://doi.org/10.1145/3319502.3374829

Tomasello, M. (2014). The ultra-social animal. *European Journal of Social Psychology, 44*(3), 187–194. https://doi.org/10.1002/ejsp.2015

Tung, F. W., & Deng, Y. S. (2007). Increasing social presence of social actors in e-learning environments: Effects of dynamic and static emoticons on children. *Displays, 28*(4–5), 174–180. https://doi.org/10.1016/j.displa.2007.06.005

Turkle, S. (2012). *Alone together: Why we expect more from technology and less from each other*. Basic Books.

Vinciarelli, A., Pantic, M., & Bourlard, H. (2009). Social signal processing: Survey of an emerging domain. *Image and Vision Computing, 27*(12), 1743–1759. https://doi.org/10.1016/j.imavis.2008.11.007

Vincent, J. (2020, January). Samsung's artificial human project definitely looks like a digital avatar. *The Verge*. https://www.theverge.com/circuitbreaker/2020/1/5/21050288/samsung-neon-artificial-human-digital-avatar-project-leaked-video

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

Walters, M. L., Koay, K. L., Syrdal, D. S., Dautenhahn, K., & Te Boekhorst, R. (2009). Preferences and perceptions of robot appearance and embodiment in human-robot interaction trials. In *Proceedings of New Frontiers in Human-Robot Interaction* (pp. 136–143).

Walther, J. B. (1996). Computer-mediated communication: Impersonal, interpersonal, and hyperpersonal interaction. *Communication Research*, *23*(1), 3–43. https://doi.org/10.1177/009365096023001001

Walther, J. B., Loh, T., & Granka, L. (2005). Let me count the ways the interchange of verbal and nonverbal cues in computer-mediated and face-to-face affinity. *Journal of Language and Social Psychology, 24*(1), 36–65. https://doi.org/10.1177/0261927X04273036

Wang, S. S., Moon, S. I., Kwon, K. H., Evans, C. A., & Stefanone, M. A. (2010). Face off: Implications of visual cues on initiating friendship on Facebook. *Computers in Human Behavior, 26*(2), 226–234. https://doi.org/10.1016/j.chb.2009.10.001

Wiltshire, T. J., Snow, S. L., Lobato, E. J. C., & Fiore, S. M. (2014). Leveraging social judgment theory to examine the relationship between social cues and signals in human-robot interactions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 58*(1), 1336–1340. https://doi.org/10.1177/1541931214581279

Won, A. S., Bailenson, J., Lee, J., & Lanier, J. (2015). Homuncular flexibility in virtual reality. *Journal of Computer-Mediated Communication, 20*(3), 241–259. https://doi.org/10.1111/jcc4.12107

Xu, K. (2019). First encounter with robot Alpha: How individual differences interact with vocal and kinetic cues in users' social responses. *New Media and Society, 21*(11–12), 2522–2547. https://doi.org/10.1177/1461444819851479

Xu, K. (2020). Language, modality, and mobile media use experiences: Social responses to smartphone cues in a task-oriented context. *Telematics and Informatics*, *48*, 101344. https://doi.org/10.1016/j.tele.2020.101344

HUMAN-MACHINE
COMMUNICATION

Human-Machine Communication
Volume 2, 2021
https://doi.org/10.30658/hmc.2.3

# Voice-Based Agents as Personified Things: Assimilation and Accommodation as Equilibration of Doubt

**Katrin Etzrodt[1] and Sven Engesser[1]**

1 Science and Technology Communication, Technical University Dresden, Dresden, Germany

## Abstract

We aim to investigate the nature of doubt regarding voice-based agents by referring to Piaget's ontological object–subject classification "thing" and "person," its associated equilibration processes, and influential factors of the situation, the user, and the agent. In two online surveys, we asked 853 and 435 participants, ranging from 17 to 65 years of age, to assess Alexa and the Google Assistant. We discovered that only some people viewed voice-based agents as mere things, whereas the majority classified them into personified things. However, their classification is fragile and depends basically on the imputation of subject-like attributes of agency and mind to the voice-based agents, increased by a dyadic using situation, previous regular interactions, a younger age, and an introverted personality of the user. We discuss these results in a broader context.

**Keywords:** artificial agents, classification, ontology, social actor, hybrids

## Introduction

> When it comes to being social, people are built to make the conservative error: When in doubt, treat it as human. (Reeves & Nass, 1996, p. 22)

In 2011, Apple launched Siri, the first commercialized voice-based agent (VBA). More VBAs have followed rapidly and have entered the habitats of people (Newman, 2018), the

**CONTACT** Katrin Etzrodt • Science and Technology Communication • Technical University Dresden • katrin.etzrodt@tu-dresden.de

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

most popular being the Google Assistant (launched in 2012), Microsoft's Cortana (2013), and Amazon's Alexa (2014). However, neither users nor scientists have yet been able to completely grasp the nature of these VBAs (Guzman, 2019). Thus, clarifying the fundamental question of 'what is it?' has become one of the main research issues within human-machine communication (A. Edwards et al., 2020). In this paper, we will explore the nature of these VBAs by investigating how they are categorized as objects, subjects, or in-between. We aim at making the classification of commercial VBAs and the associated processes of "equilibration" (Piaget, 1970/1974) accessible for empirical research.

Both CASA and Social Presence Theory argue that VBAs are what Reeves & Nass (1996, p. 22) describe as objects of doubt. Research under the CASA paradigm (Nass et al., 1994), postulating that computers are social actors, has provided multifaceted evidence that people exhibit diverse social reactions toward various technological artifacts such as traditional media, computers, avatars, or robots. Social Presence Theory (Short et al., 1976) assumes that these reactions are caused by the perception of mediated social entities as being present (Lombard et al., 2017), or by the failure to recognize those entities as artificial (K. M. Lee, 2004), and specifically as "non-human" (Latour, 2005). Although even marginal social cues can trigger social reactions and the feeling of presence (e.g., Reeves & Nass, 1996), anthropological similarity can further enhance them (e.g., Appel et al., 2012; Horstmann et al., 2018). In particular, voices are notably powerful indicators of social presence (Reeves & Nass, 1996) and being able to talk can be an indicator for being alive (Turkle, 1984/2005, p. 48).

Although a large body of research in the past 30 years has investigated the consequences of this doubt (e.g., the activation of social heuristics), we know very little about the *doubt itself*. It is still unclear toward "whom" or "what" (Gunkel, 2020, p. 54) people react when they are interacting with (artificially intelligent) machines. The various terminologies used to describe them, such as "evocative objects" (Turkle, 1984/2005), "quasi-objects" or "hybrids" (Latour, 1991/1995), "epistemic things" (Knorr-Cetina, 2001), "non-humans" (Latour, 2005), "subject objects" (Suchman, 2011), or "social things" (Guzman, 2015) illustrate the difficulty of capturing their essence. Although definitions differ considerably, common to all terms is the reference to ambiguous entities possessing both object qualities and subject qualities, whose unique combination creates new qualities beyond the sum of its parts (Roßler, 2008). In the paper at hand, we aim at investigating the categorization of these doubtful objects.

For this purpose, we will argue that the doubt arises from an irritated ontological object–subject classification by referring to the epistemologist Piaget (1970/1974). We will outline how people resolve the irritation through specific *equilibration* processes by assigning the irritating object to an existing scheme or modifying these schemes gradually.

The second part of the paper is dedicated to influences on this classification. Research on artificial agents indicates that attributes of the *situation* (A. Edwards et al., 2019; Purington et al., 2017), of the *users* (Epley et al., 2007; E. J. Lee et al., 2000; Woods et al., 2007), and of the *agents* themselves (e.g., Nass & Moon, 2000; Reeves & Nass, 1996) may be relevant.

Based on an empirical approach, we designed a scale for the classification, assessed situational, individual, and technological influences in an online survey, and validated the findings with a second sample. By investigating the degree to which people classify a VBA as social actor, ranging between thing and person, and the identification of influences we intend to advance the concept of human–machine communication and contribute to the CASA paradigm.

## Classification of Voice-Based Agents

Voice-based agents like Siri, Alexa, and the Google Assistant are a subtype of artificial social agents, with an operating system based on artificial intelligence and natural language processing using a disembodied voice emanating from a device (e.g., smartphone, loudspeaker box) to communicate with the users and execute their tasks. They are still a reasonably new technology, meaning that people lack exhaustive previous experience with this unique conversational interface (Guzman, 2015; Krummheuer, 2015). However, there are signs for human-machine agent interaction scripts (Gambino et al., 2020), suggesting a gradual object–subject classification of artificial agents. Thus, we ask to which degree people classify VBAs into the object–subject classification (RQ1).

### Voice-Based Agents as Objects of Doubt

Referring to Piaget's studies on epistemology (1970/1974), we assume the most fundamental way of gaining knowledge about a new object is figuring out if it is part of the "psychomorph" or the "physicomorph" scheme, which are diametrical poles of the same ontological classification. Turkle (1984/2005, p. 34) referred to these poles as "psychological" and "physical." Gunkel (2020, p. 54) arrived at a similar conclusion by referring to Derrida's distinction of "who" and "what." The *psychomorph scheme* ("who") is defined by subjects, which are living beings, equipped with capacities like thinking or feeling, and the potential of agency (Piaget, 1970/1974, p. 48). Thus, it is an analog to the scheme "other persons" (Gunkel, 2020, p. 54), and used "to understand people and animals" (Turkle, 1984/2005, p. 34). In contrast, the *physicomorph scheme* ("what") is defined by inorganic, non-living objects, which are sufficiently comprehensible in terms of precise, logical-mathematical categories, and deterministic causal laws (Geser, 1989, p. 233). That is, it is an analog to the scheme "things [that] are mere instruments or tools" (Gunkel, 2020, p. 54), and "used to understand things" (Turkle, 1984/2005, p. 34). While the *physicomorph* scheme (hereinafter referred to as thing theme) results from empirical experience (e.g., physical perception and movement), the *psychomorph* scheme (hereinafter referred to as person theme) originates in the introspective experience of a subjectivity (Piaget, 1970/1974). This theoretical approach is related to the impossible verification of subjectivity (Gunkel, 2020; Turing, 1950) or agency (Krummheuer, 2015), as well as to the imputation of mental states to objects (Premack & Woodruff, 1978) and the gradual assignment of personhood (Hubbard, 2011), as discussed prior in Etzrodt (2021).

Despite the growing ability to distinguish between psychomorph and physicomorph due to the individual's development and its constant confrontation with the environment (Piaget, 1970/1974; Turkle, 1984/2005), some objects remain a challenge. Between *things* and *persons,* there exists a wide range of objects (e.g., plants, animals, or artificial entities) that can be regarded as objects of doubt. Recent empirical research confirms the doubtful nature of various artificial social agents: Guzman (2015) noted in her qualitative interviews a constantly shifting use of the pronouns "she" and "it" when people talked about the VBA

Siri. It seemed that Guzman's respondents were torn between the association of a person and a thing, which she traced back to the interviewees' focus of attention. If their "attention turns away from Siri the voice or Siri the image to Siri the program, Siri again becomes a thing" (Guzman, 2015, p. 195); thus, she concludes, people "recognize certain characteristics of humans and machines within them" (p. 257). A. Edwards (2018) found a similar inconsistency, when she asked her participants to choose two out of the three entities (social robot, human, and primate), that had more in common in relation to the third. Although, participants combined humans and primates in opposition to social robots, based on robots being artificial and non-living things, some coupled robots and humans in reference to their (assumed) resemblance in embodiment, intellect, and behavior, and the capability to interact socially through talking and understanding.

## Voice-Based Agents as Objects of Equilibration

Once the classification of an object is in doubt, the irritation has to be resolved. People need to decide if machines are "mere things . . . or someone who matters . . . or something altogether different that is situated in between the one and the other" (Gunkel, 2020, p. 56). Piaget (1970/1974) refers to this as *equilibration*—a balancing and self-regulating process, which is achieved through *assimilation* or *accommodation* (Figure 1).

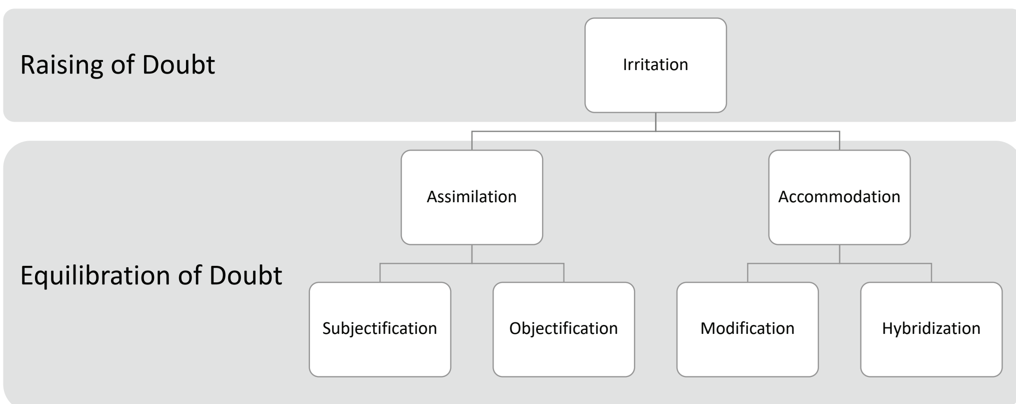



**FIGURE 1 Equilibration of Doubt**
**(Inspired by Piaget, 1970/1974 and Geser, 1989)**

*Assimilation* is the assignment of an irritating object, such as the VBA, to an existing scheme: the thing scheme *(objectification)* or the person scheme *(subjectification)*. As a result, people overestimate either the VBA's objecthood or subjecthood. Particularly for Siri, Guzman (2015) found that some people explicitly described it as an entity, while others viewed it as a device. Similar associations were found by Purington and colleagues (2017) in the user comments about Alexa (Amazon Echo). They demonstrated that, although the objectifying pronoun "it" was used by the majority of the authors, some favored the personifying pronoun "she." The preference of objectification is confirmed for social robots by A. Edwards (2018), where more than half of the respondents regarded social robots as things in contrast to living and natural subjects such as humans or apes. Therefore, we formulate the hypothesis (H1): VBAs are assimilated more often to the thing scheme than to the person scheme.

*Accommodation* refers either to the modification of an existing scheme or to the creation of a new one (Piaget, 1970/1974). Regarding the *modification* of the object–subject classification, the thing scheme can be modified by adding person attributes or the person scheme can be enriched with thing attributes. Apart from that, people may build a hybrid scheme, featuring a more or less balanced combination of attributes from things and persons *(hybridization)*. While modification draws on existing heuristics and changes them slightly, hybridization requires the active acquisition and construction of completely new heuristics.

Research on VBAs implies accommodation: In addition to the "spectrum from fully human to fully machine," Guzman (2015, p. 227) identified an "overlap in the middle" concerning the ontology of Siri, as a result of a reconfigured "understanding of humans and machines to the degree that we now share characteristics" (p. 257). Purington and colleagues (2017) found a mixed use of the pronouns "she" and "it" for Alexa in the same user comment. Furthermore, some user reactions suggest accommodation through the simultaneous activation of social and non-social scripts, such as inappropriate, rude, or insensitive social behavior toward artificial agents: For example, people abuse social robots (Broadbent, 2017), and direct bullying and sexual harassment toward VBAs like Alexa (Cercas Curry & Rieser, 2018).

## Influences on the Classification of Voice-Based Agents

Although social reactions toward computers are fundamentally human, exist in all groups, and occur even in the case of weak social cues (Reeves & Nass, 1996), there is evidence that object–subject classification varies due to factors at the levels of situation, user, and agent. Thus, we ask to what extent do factors at the levels of the situation, the user, and the agent influence the object–subject classification of VBAs (RQ2)?

### Attributes of the Situation

The ontological classification of machines can differ between various situational interactions and social contexts. A. Edwards and colleagues (2019) found that a positive expectancy violation during an initial interaction with a social robot reinforces the feeling of social presence and reduces uncertainty. Leite and colleagues (2013) suggested a change in relationship through prolonged interaction. Furthermore, personification increased when Alexa was embedded in *multi-person households* such as families (Lopatovska & Williams, 2018; Purington et al., 2017). Against this backdrop, we assume that previous interactions with VBAs affect the classification (H2) and its use in the presence of others increases the classification's preference for the person scheme (H3).

### Attributes of the User

*Age* appears to be a major influence on classification. Children are known to attribute subject status to objects in general (Piaget, 1970/1974) and artificial agents (Epley et al., 2007;

Turkle, 1984/2005) more strongly than adults do. Apart from that, users' attributes seem to affect the perceived attributes of the agents rather than their classification. In this context, age in general affects perceived similarities between a participant's and a robot's personality traits (Woods et al., 2007). In contrast, the impact of *gender* appears inconsistent. Gender neither affects the perception of a social presence for synthesized voices (K. M. Lee & Nass, 2005), nor the evaluation of flattery behavior of the VBA (E. J. Lee, 2008). However, the attribution of personality to robots may be gender-specific (Woods et al., 2007); and matching genders of the user and the VBA alters social reactions (E. J. Lee et al., 2000). Besides, *personality traits* of the user have been crucial for research, reflecting differences between individuals as well as parallels in interaction in general. Although research did not find effects of the user's personality on the perception of social presence, matching personalities of user and VBA may influence perceived characteristics—regarding the trait "extravert–introvert" (Nass & Lee, 2001), and similarity—regarding the trait "neuroticism–emotional stability" (Woods et al., 2007). Closely related to personality is a person's *affinity for technology* (Attig et al., 2017), which is particularly noticeable regarding the novelty of VBAs.

## Attributes of the Agent

The VBA's conversational mode exhibits subject-likeness, involving expressed effective and meaningful behavior and the imputation of agency and mind. However, it is still unclear to what extent they are relevant for the object–subject classification.

**Agency as expression of effective behavior.** Due to the recently increased capabilities of machines to mimic natural human behavior their genuine agency is conveyed more strongly (Guzman, 2018). VBAs can directly answer users, communicate with and control other smart objects in their environment (e.g., switching lights on and off), collect and present information from the internet, activate apps, or initiate purchases. Thus, in the context of VBAs, the term agency refers to an *effect*, originating from interaction with the environment and other beings, which may be interpreted as *behavior similar* to humans (e.g., A. Edwards, 2018). Within the framework of social interaction theory, this effect refers to the ability of *interdependence* (Simmel, 1908) and *orientation* (Weber, 1922) toward the behavior of others—thus, they are affecting others and are affected by others. The most basic indicator for interdependence between VBAs and users is the VBAs' expression of *receptiveness* to vocal commands. For instance, Alexa lights up a blue ring on the Amazon Echo and a blue line on the Amazon Echo Show to indicate it is "listening." To demonstrate the ability of orientation the VBA needs to express *attentiveness* in the first place (Biocca et al., 2003). Alexa does this by additionally turning a brighter light toward the direction of the sound source in most Echo devices. The feeling of receptivity increases the perception of sociality—including that of computers (Burgoon et al., 1999) and agency in general (Appel et al., 2012; Guadagno et al., 2007). However, it is unclear to what degree the VBA's classification is altered if their behavior is perceived as effective.

**Mind as expression of meaningful behavior.** The human-like voice and the use of language are expressions of meaningful behavior, which is closely linked to consciousness (Reichertz, 2014), and a theory of mind (Epley et al., 2007; Premack & Woodruff, 1978). Thus, the language- and voice-based subjectivity transcends the ability to interact effectively by adding meaning to its actions. Meaningful behavior is crucial for orientation in any social interaction and based on mutual understanding, on similarity in thinking and/or feeling (to some degree), and the assumption of intentional actions. That is, before an orientation toward the actions of others is possible, these actions must be *understood* based on one's own and the anticipated other's experiences within a shared (natural or social) world (Schütz, 1974). In particular, the conversational mode of VBAs incorporates such specific assumptions of *similarity in thinking and feeling* to a certain degree: First, the use of words of a certain language, and that VBAs understand these words, refers to a specific shared social world between users and VBAs. Second, by referring to themselves as "I," VBAs suggest that they are person-like, self-conscious entities, which operate according to the same rules as their human users.

Furthermore, the meaning of actions is closely related to the assumption of *intentionality*, in contrast to accidental actions or for purely physical reasons (Simmel, 1908). The simplest complex of intentionality that can be attributed to an action are motives (Schütz, 1974). According to Schütz, it is sufficient to put oneself in a typical position with typical motives and plans. *Personality* constitutes the origin of these typical motives, attitudes, and mindsets of persons as subjects. It contains both the assumption about behavior that is based on typical human nature, typical decisions, reactions, or feelings and on ways in which individuals differ from each other (Buss, 2008). Therefore, a perceived personality *suggests* a subject who chose to act in a specific manner, but theoretically could have responded in a different way (Higgins & Scholer, 2008). As a result, if personality is attributed to a VBA its actions may be no longer viewed as random—regardless of whether they were programmed to imitate intention or whether they are behaving intentionally by nature. Research confirms the attribution of personality traits to synthetic voices (e.g., C. Edwards et al., 2019; Ray, 1986) and very specific personalities to the commercial VBAs Alexa and the Google Assistant (Garcia et al., 2018). However, it is uncertain whether the quantity or quality of perceived personality traits of the VBA alter the object–subject classification. Previous CASA research has concentrated on the quality such as the personality trait "extraversion" in interpersonal interactions and uncovered extraverted voices are generally perceived as more socially present (e.g., K. M. Lee, 2004; Nass & Lee, 2001; Nass & Moon, 2000). Although extraversion may be relevant for social reactions toward VBAs, it is unanswered if extraversion or other personality traits or that personality traits are attributed in the first place are relevant for their classification.

## Research Questions and Hypotheses

An overview of the relevant variables, research questions, and hypothesized influences can be found in the theoretical model (see Figure 2).

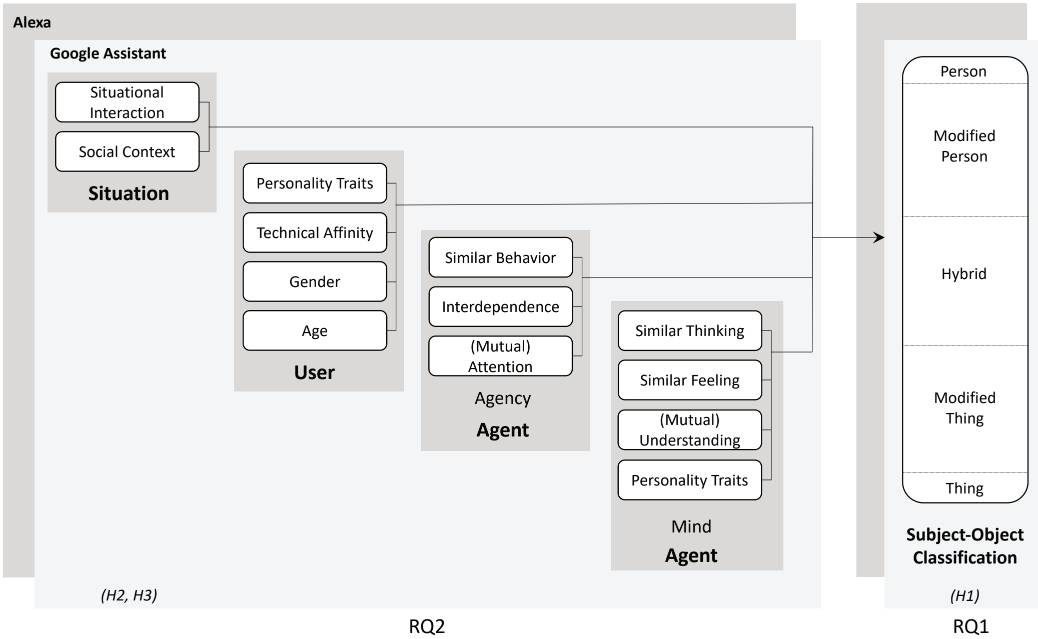



**FIGURE 2 Theoretical Factor Model**

# Method

## Sample

In late 2018, we conducted an online survey with a demonstrational design (N = 853) among all students of a large German university, recruited via the university's student mailing list (response rate of 2.6%) and randomly assigned to either the VBA Alexa (Amazon Echo) or the Google Assistant (Google Home). Participants had a mean age of 24, ranging from 17 to 50 years, 52% of whom were male, 76% were undergraduates, and 23% graduates. The sample was (on a 6-point scale) above average creative (M = 3.67, SD = 1.00), conscientious (M = 3.46, SD = 0.86), and emotionally stable (M = 3.2, SD = 0.93), as well as moderately agreeable (M = 3.1, SD = 0.86) and extraverted (M = 3.1, SD = 1.09). It expressed an affinity for technology above average (M = 4.01, SD = 1.20, 6-point scale).

Most of the participants already knew the names Alexa (96%), and the Google Assistant (71%) in general, and 84% knew their assigned VBA. Although one third owned the assigned Google Assistant (33%), only some possessed Alexa (7%). The primary sources of knowing Alexa were indirect ones: advertisement (43%) and contact through other people (24%), which however, were also important for Google Assistant (17% advertisement, 12% other people). Other indirect sources were non-fictive media (Alexa: 16%, Google Assistant: 7%) and rarely fictive media (Alexa: 6%, Google Assistant: 1%). If the participants had used their assigned VBA prior to the survey's demonstrational interaction, most of them had used it primarily alone (60%) and moderately frequent (M = 3.3, SD = 1.45, 5-point scale).

According to its average age the student sample belonged to a cohort widely labelled as Generation Z, which differs from the previous cohort (commonly labeled as Millennials) in political, economic, and social factors (Dimock, 2019; Singh, 2014). Regarding these differences of generations and the classification's development during aging, the study was repeated with an older sample (Millennials)—all employees of the same university, recruited via the university's staff mailing list (response rate of 6.2%)—to validate the findings and specify cross-generational effects. The staff sample (N = 435) was, on average, 10 years older (with a mean age of 33, from 18 to 65 years), 65% were graduates, 26% were undergraduates, and 3% had a doctoral degree. Participants were slightly more conscientious (M = 3.71, SD = 0.78) and had less affinity for technology (M = 3.96, SD = 1.20), however, their prior experiences with the assigned VBA resembled those of the student sample.

## Procedure

Although both assistants' German voices were female, they differed in their characters, the way they were advertised, and their manufacturing companies' image. To distinguish between possible variations caused by the mentioned differences and general classifications of VBAs, one group of participants assessed Alexa, another the Google Assistant. To get impressions as close as possible to the true perception of the two VBAs, we chose a demonstrational design by simulating interactions with pre-recorded videos with the original answers of the voice-based loudspeaker variants of the Google Assistant or Alexa to predefined questions in German (Table A1 in the study's OSF repository). Before the simulated interaction, participants reported their previous experiences with various VBAs (including the assigned VBA) and typical usage situations. During the simulated interaction, they activated four videos of the VBA's answer one after the other by clicking on the question. After the interaction, they classified the VBA and assessed perceived attributes.

Questions for the VBAs were selected that had been advertised previously by Amazon or Google as preferable interaction features (e.g., in commercials or on the website), and that had the potential to exhibit personality characteristics of the VBA. If a VBA provided multiple answers to the same question, we randomly selected one. Because people form their impressions within the first few seconds of contact with a voice (Ray, 1986), the video sequence for each simulated interaction had a duration between 7 and 17 seconds. The videos can be obtained from this study's OSF repository.

## Measures

**Object–Subject Classification.** To examine the object–subject classification, we drew on the diametrical relation of the thing scheme and the person scheme described above and asked participants: "What would you say, is Alexa [or the Google Assistant] rather like a thing (object) to you or rather like a person (subject)?" Because "person" refers to the status "personhood," it can be assumed with Hubbard (2011) that it is a gradual assignment, whose highest degree is represented by the term "person." To address the continuum between the schemes we used a *100-point scale*, consisting of the two poles *"thing (object)"* and *"person (subject)."*

As discussed in Etzrodt (2021), the broad-scale allowed intuitive answers—independent of the participant's ability to verbalize the classification (Turkle, 1984/2005, p. 48). It also allowed to detect minor forms of accommodation and to distinguish between modification and hybridization (see Figure 3). As described by the author, classification as the result of assimilation into the thing or the person scheme refers to the absence of any previous accommodation, indicated on the scale by the ratings of 1 or 101. Thus, VBAs are added to the existing scheme (thing or person), but the scheme itself does neither get in conflict with the other nor does it change. Classification as a result of accommodation depends on pre-existing structures (Piaget, 1974, p. 34), referring to a change or reaffirmation of the categories' borders (Turkle, 2005). Hence, Etzrodt (2021) concludes, the accommodated classification is measured by a weak or strong merging of the thing and the person scheme, implied if people were distancing from one of the poles on the scale. A weak merging is represented by ratings close to one of the schemes (2–33 and 67–100), indicating the modification of a dominant scheme by implementing elements of the other. A strong merging is represented by ratings near the scale's center (34–66), indicating a hybridization with a more or less balanced reunion of both schemes. To determine how sophisticated the equilibration process was, we asked how *confident* participants were in their classification on a 5-point scale (Etzrodt, 2021).

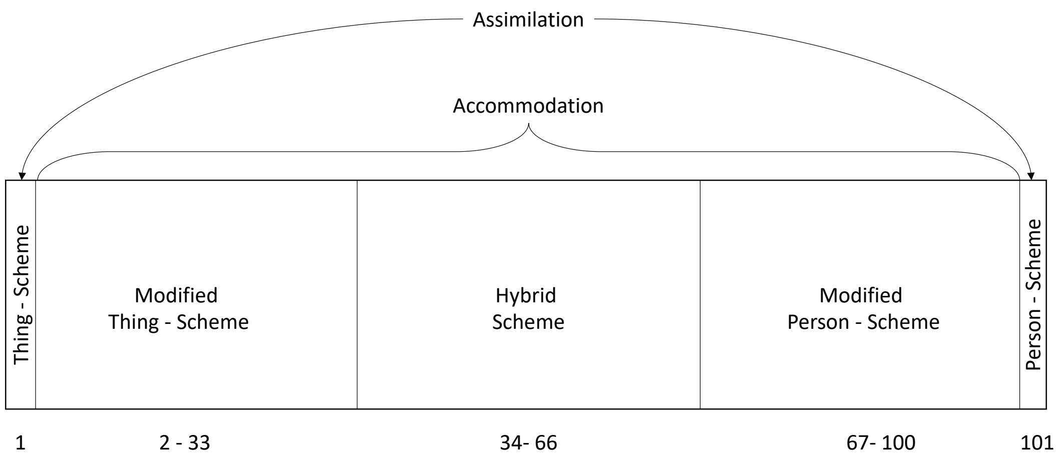



**FIGURE 3 Equilibration on the Object–Subject Classification Scale (1 to 101), Etzrodt (2021)**

**Attributes of the situation** were measured as previous knowledge of and interactions with the assigned VBA and the social contexts in which these interactions are usually embedded. Previous *knowledge* assessed if the participants had ever heard about or knew any VBA—independent of their assigned VBA. *Previous interactions* differentiate between participants who had contact with their assigned VBA for the first time through our survey or knew this VBA solely from fiction or non-fiction media, advertising, or had seen others using it, but had none or only minimal prior interactions, and those who had continuous previous interactions through ownership. Therefore, the first group can be assumed to have had none or only few equilibration processes before the study. In contrast, the latter group was likely to have had undergone this process several times. In addition, we considered the frequency of use on a 5-level rating scale from "very occasionally" to "very often." We assessed the *social contexts* of use by asking if participants usually used the VBA in multi-person contexts (family, friends, or acquaintances) or dyad contexts (the absence of other people).

**Attributes of the user** were assessed by the user's personality, affinity for technology, and demographics such as age and gender. *Personality* was measured with the "Big Five Inventory" short scale using the original 5-level rating scale (Rammstedt et al., 2013). Principal component analysis (PCA) confirmed the factors agreeableness, conscientiousness, emotional stability, extraversion, culture (creativity) ($\chi^2(45) = 1451.02$, $p < .001$, KMO = .60, most factor loadings and h2 > .60). "Agreeableness" exhibited the poorest performance and was interpreted with caution. *Affinity for Technology* was measured with nine items of the German ATI Scale (Franke et al., 2018), indexed into one component via PCA ($\chi^2(36) = 5375.01$, $p < .001$, KMO = .91, Cronbach's alpha = .92, factor loadings .63–.88; h2 > .60). The PCAs in the staff sample confirmed these factors (see OSF).

**Attributes of the Agent. Agency** was conceptualized as an assigned capability to act, divided into the three dimensions: *similarity to the behavior* of the respondent ("Does not behave like me—behaves like me," 7-level rating scale), attributed *attention* of the respective VBA, and the feeling of *interdependence*. Subsequent operationalizations were realized either for Alexa or Google Assistant. For easier reading, only the Alexa variant is provided in the examples. The two latter dimensions were measured with five items based on the reduced relational communication scale by Ramirez (2009), using the original 6-level rating scale, but formulated for a hypothetical situation ("Imagine you and Alexa are having a conversation . . . "). However, the confirmatory factor analysis (CFA) indicated that these items loaded on the dimensions *attention* ("How attentive is Alexa to the interaction?", "How strongly is Alexa involved in an interaction?"), and *interdependence* ("How much is Alexa adapting to the interaction?", "How ready is Alexa to have an interaction with you?", "How strongly does Alexa respond to your comments or questions?") suggested by Biocca and colleagues (2003). To assess convergent validity, the standardized factor loadings, average variance extracted (AVE), and reliabilities (omega and Cronbach's alpha) were examined. The CFA showed both an excellent model fit for the dimensions ($\chi^2(4) = 8.987$, GFI = .994, CFI = .995, TLI = .989, RMSEA = .044) and a moderate convergent validity (AVE >.51, omega > .68, Cronbach's alpha > .66), confirmed by the staff sample (see OSF).

**Attributes of the Agent. Mind** was operationalized as the VBA's *similarity in thinking* and *feeling* (e.g., "Thinks like me–does not think like me," 7-level rating scale), the VBA's *ability to understand* its user ("How well does Alexa understand you?", 7-level rating scale), and attributed *personality traits*. The modified Minimal Redundancy Scale based on Lang (2008) was used to measure the VBA's personality on a semantic differential using the original 6-level rating scale, indexed to the Big Five factors. However, not all human personality traits worked for the VBAs. CFA identified the items warmhearted, selfless, over-accurate, confident, self-contented, open, loving company, and inventive as insufficient. After their removal the model produced an excellent fit ($\chi^2(80) = 227.330$, GFI = .964, CFI = .964, TLI = .950, RMSEA = .046), a good reliability with most factor loadings larger than 0.7, and a moderate convergent validity (AVE > 0.40, omega > 0.50), confirmed by the staff sample (see OSF). As a result, the factor *culture* includes items exclusively referring to creativity (creative, imaginable, artistic) and was interpreted accordingly. In line with previous research (Garcia et al., 2018; Guzman, 2020), participants had problems assigning emotions to the VBA explaining why the factor *emotional stability* displays the poorest performance.

The factor was maintained for comparison with subsequent studies but was interpreted cautiously.

## Results

Data was analyzed using R. In particular, the VBAs' classification and attributes showed non-normal, mainly positively skewed and heavy-tailed distributions. Thus, robust tests (packages WRS2, and robustbase) were used to control the results of common statistical tests. The stepwise robust regression found several outliers (with weights = 0 or ~= 1). Although, the significance of the predictors was equal, the estimates differed slightly between OLS and robust regression. Thus, to avoid inaccuracy, we report the estimates of the robust regression. Reported results apply to the student sample and will be validated by the staff sample. The supplemental tables and figures can be found at this study's OSF repository.

## Equilibration of the Object–Subject Classification

Almost one out of three participants assimilated VBAs into an existing scheme, while more than two out of three had accommodated their schemes (RQ1, Figure 4).

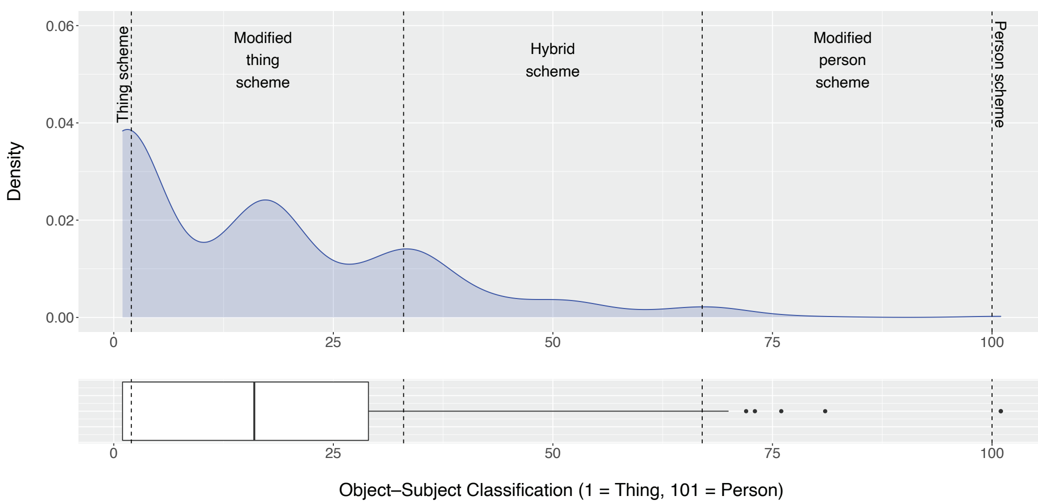



**FIGURE 4 Density Plot and Box Plot, of the Equilibration on the Object-Subject Scale, Student Sample**

As predicted (H1), in case people did *assimilate*, they almost always objectified VBAs (Table 1). In contrast, apart from two people, the VBAs were not subjectified at all. In case people *accommodated*, most of them classified the VBAs into modified schemes primarily with respect to the thing scheme (49%, Table 1), whereas only a minimal number of people (2%) modified the person scheme. Thus, VBAs were mainly classified as things supplemented by aspects of a person. However, 17% classified the VBA into a hybridized scheme.

**TABLE 1 Equilibration of the VBA**

| | | Student Sample | | Staff Sample | |
|---|---|---|---|---|---|
| | | *n* | % | *n* | % |
| Assimilation | Thing scheme | 263 | 31.0 | 141 | 32.7 |
| Accommodation | Modified thing scheme | 419 | 49.4 | 208 | 48.3 |
| | Hybrid scheme | 148 | 17.4 | 67 | 15.6 |
| | Modified person scheme | 17 | 2.0 | 15 | 3.5 |
| Assimilation | Person scheme | 2 | 0.2 | 0 | 0.0 |
| Note: On a scale from 1 to 101, 1 = thing, 2 to 33 = modified thing scheme, 34 to 66 = hybrid scheme, 67 to 100 = modified person scheme, 101 = person scheme | | | | | |

Although most of the participants were quite certain with their classification (M = 4.34, SD = 0.89), the further they moved away from the thing scheme, the more uncertain they became, $r = -.54$, $t(842) = -18.56$, $p < .001$. However, the LOESS graph (Figure 5) indicates that the uncertainty increases if the classification moves away from *any* existing scheme. Although very few people subjectified VBAs, they did it with the same level of certainty as those who objectified them. Consequently, modification and hybridization exhibited uncertainty.

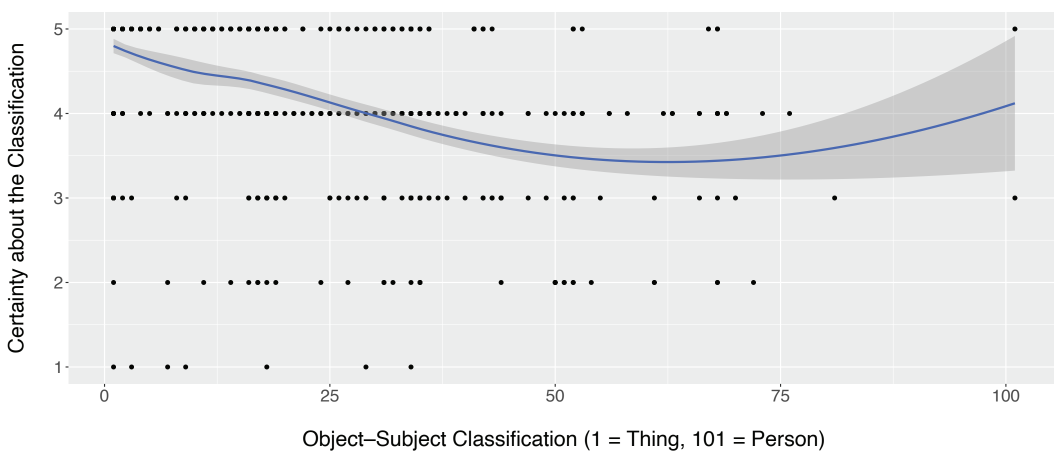



**FIGURE 5 Classification Correlated With Certainty About the Classification, Student Sample**

## Influences on the Object–Subject Classification

To identify relevant influences on the classification (RQ2) we conducted a stepwise robust regression (SRR). The SRR indicated that more participants classified Alexa further away from the pole thing than Google Assistant if the previous experienced situations were held constant (Table 2, Model 2). However, the explained variance was low, and including the VBAs' agency eliminated this effect (Table 2, Models 4).

**TABLE 2 Stepwise robust regression (M-estimators) using object–subject classification as the criterion, student sample**

| | Model 1 (n = 849) | | | Model 2 (n = 849) | | | Model 3 (n = 831) | | | Model 4 (n = 683) | | | Model 5 (n = 535) | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *B* | *[CI]* | *β* | *B* | *[CI]* | *β* | *B* | *[CI]* | *β* | *B* | *[CI]* | *β* | *B* | *[CI]* | *β* |
| (Intercept) | 14.47*** | [13.00, 15.95] | | 12.01*** | [9.33, 14.69] | | 12.91*** | [9.90, 15.94] | | 14.41*** | [11.08, 17.76] | | 17.64*** | [13.58, 21.70] | |
| **VBA** | | | | | | | | | | | | | | | |
| Alexa a) | 1.53 | [-0.57, 3.63] | .09 | 2.82* | [0.34, 5.29] | .16 | 3.44** | [0.94, 5.93] | .20 | 1.30 | [-1.36, 3.96] | .07 | 1.66 | [-1.44, 4.77] | .09 |
| **Previous Situation** | | | | | | | | | | | | | | | |
| Dyad b) | | | | 2.79* | [0.30, 5.28] | .16 | 2.38+ | [-0.14, 4.90] | .14 | 1.65 | [-1.00, 4.30] | .09 | 0.27 | [-2.63, 3.16] | .02 |
| Frequency of use | | | | -0.29 | [-1.41, 0.83] | -.02 | -0.48 | [-1.60, 0.65] | -.03 | 0.06 | [-1.14, 1.27] | .00 | 0.10 | [-1.28, 1.49] | .01 |
| Knows VBA in general c) | | | | 0.06 | [-3.19, 3.31] | -.00 | -0.99 | [-4.28, 2.29] | -.06 | 0.78 | [-2.78, 4.34] | .04 | -0.39 | [-4.55, 3.77] | -.02 |
| Owns the VBA d) | | | | 4.66** | [1.58, 7.73] | .27 | 4.34** | [1.23, 7.46] | .25 | 1.47 | [-1.80, 4.74] | .08 | 0.47 | [-3.26, 4.20] | .03 |
| **User** | | | | | | | | | | | | | | | |
| Agreeableness | | | | | | | 1.55** | [0.46, 2.64] | .09 | 0.93 | [-0.25, 2.10] | .05 | 0.99 | [-0.36, 2.34] | .06 |
| Conscientiousness | | | | | | | -0.17 | [-1.27, 0.92] | -.01 | 0.15 | [-1.02, 1.32] | .01 | 0.22 | [-1.56, 1.12] | -.01 |
| Emotional stability | | | | | | | -1.19* | [-2.32, -0.06] | -.07 | -0.39 | [-1.60, 0.82] | -.02 | -0.48 | [-1.84, 0.87] | -.03 |
| Extraversion | | | | | | | -1.24* | [-2.36, -0.12] | -.07 | -1.09 | [-2.28, 0.10] | -.06 | -0.45 | [-1.81, 0.91] | -.03 |
| Culture (creativity) | | | | | | | -0.32 | [-1.40, 0.76] | -.02 | 0.11 | [-1.04, 1.26] | .01 | 0.23 | [-1.07, 1.54] | .01 |
| Affinity for technology | | | | | | | 1.42* | [0.16, 2.69] | .08 | 0.52 | [-0.84, 1.88] | .03 | 0.19 | [-1.36, 1.75] | .01 |
| Women e) | | | | | | | 0.12 | [-2.47, 2.71] | .01 | 0.59 | [-2.16, 3.34] | .03 | -0.64 | [-3.77, 2.50] | -.04 |
| Age | | | | | | | -1.96*** | [-3.03, -0.90] | -.11 | -0.61 | [-1.85, 0.62] | -.03 | -0.25 | [-1.68, 1.17] | -.01 |
| **Agent: Agency** | | | | | | | | | | | | | | | |
| Similar behaving | | | | | | | | | | 5.16*** | [3.98, 6.35] | .29 | 3.04*** | [1.47, 4.61] | .17 |
| Interdependence | | | | | | | | | | 0.88 | [-0.66, 2.42] | .05 | 0.26 | [-1.66, 2.17] | .01 |
| VBA's attention | | | | | | | | | | 3.45*** | [1.95, 4.94] | .20 | 2.84** | [1.09, 4.59] | .16 |
| **Agent: Mind** | | | | | | | | | | | | | | | |
| Similar thinking | | | | | | | | | | | | | 4.01*** | [2.20, 5.82] | .23 |
| Similar feeling | | | | | | | | | | | | | 1.95* | [0.45, 3.45] | .11 |
| VBA's understanding | | | | | | | | | | | | | 1.90** | [0.47, 3.33] | .11 |
| Personality traits (count) | | | | | | | | | | | | | 1.71* | [0.18, 3.25] | .10 |
| Agreeableness | | | | | | | | | | | | | -0.50 | [-1.96, 0.95] | -.03 |
| Conscientiousness | | | | | | | | | | | | | -0.55 | [-2.03, 0.93] | -.03 |
| Emotional stability | | | | | | | | | | | | | -0.14 | [-1.55, 1.28] | -.01 |
| Extraversion | | | | | | | | | | | | | -0.24 | [-1.67, 1.18] | -.01 |
| Culture (creative) | | | | | | | | | | | | | 1.96* | [0.43, 3.40] | .11 |
| **Interactions** | | | | | | | | | | | | | | | |
| Attention * interdepend. | | | | | | | | | | -0.13 | [-1.19, 0.92] | -.01 | -0.48 | [-1.79, 0.84] | -.03 |
| Sim. behaving * thinking | | | | | | | | | | | | | -0.82 | [-2.18, 0.54] | -.05 |
| Sim. behaving * feeling | | | | | | | | | | | | | 1.42+ | [-0.03, 2.87] | .08 |
| Sim. thinking * feeling | | | | | | | | | | | | | -1.34+ | [-2.81, 0.13] | -.08 |
| *$R^2$* | | *.002* | | | *.02* | | | *.07* | | | *.22* | | | *.34* | |
| *Change in $R^2$* | | *.002* | | | *.02* | | | *.05* | | | *.15* | | | *.12* | |

*Note.* + indicates p < .1 * indicates *p* < .05. ** indicates *p* < .01. *** indicates p < .001. Dichotomic variables are compared with following omitted (reference) level: a) Google Assistant, b) Multi-Person Interaction, c) Does not know any VBAs, d) Knows interaction with VBA secondhand, e) Men and diverse

**Attributes of the Situation.** As predicted in H2, previous interactions affected the classification (Model 2, Table 2). If people *owned* the VBA they classified it more distanced from the pole thing than people who presumably equilibrated their classification for the first time. Contrary to our assumption (H3), the *absence of other people* increased the distancing from the mere thing scheme. However, the explained variance was still less than 1%, and the effects disappeared after including the VBAs' agency (Table 2, Model 4).

**Attributes of the User**. As *age* increased, the classification tended toward the thing scheme (Table 2, Model 3). In contrast, people with more *affinity for technology* tended to distance from the mere thing scheme (Table 2, Model 3). In line with K. M. Lee & Nass (2005) the *gender* of the user was not significant at all for the classification. However, *personality traits* were. *Agreeable* users were more inclined to classify the VBA distanced from the thing pole, while more *emotionally stable* and *extraverted* users exhibited objectification tendencies. Nevertheless, the explanatory power of the model remained small and the inclusion of the VBAs' attributes eliminated most effects. Whereas agency negated the effects of age, affinity for technology, emotional stability, and agreeableness (Table 2, Model 4) assumed mind negated the effects of extraversion (Table 2, Model 5).

**Attributes of the Agent.** Perceived *agency* contributed significantly to a classification distanced from the mere thing scheme and increased the explained variance to 22% (Table 2, Model 4). The VBAs' attentiveness (M = 3.3, SD = 1.29) and interdependence (M = 3.9, SD = 1.23) were rated moderately high, whereas their behavior was not perceived as very similar (M = 2.1, SD = 1.64). However, only increasing *attentiveness* and *similar behavior* increased a distanced classification from the pole thing. The effects held when mind attributes were added but weakened substantially.

Perceived *mind* increased the explained variance to 34% (Table 2, Model 5). Although the similarity of mind was rated low, the VBAs' thinking (M = 2.12, SD = 1.61) was assumed to be more similar than their feeling (M = 1.60, SD = 1.61). Consistently, distancing of the thing scheme was predicted to a higher degree by *similarity in thinking* than in *feeling* (Table 2, Model 5). In contrast, the VBAs' ability to understand the user was rated moderately high (M = 3.64, SD = 1.67) and also affected the distancing from the thing scheme to a similar amount. Participants assigned on average 16 out of 25 personality items to the VBAs (M = 16.5, SD = 10.2). They were perceived as very conscientious (M = 4.9, SD = 0.86), emotionally stable (M = 4.7, SD = 0.97), agreeable (M = 4.6, SD = 1.00), and extraverted (M = 4.5, SD = 1.09) but less creative (M = 3.0, SD = 1.33). Alexa was perceived as slightly more conscientious (M = 5.1, SD = 0.77) than Google Assistant (M = 4.7, SD = 0.91), $t(644.34) = 5.803$, $p < .001$. However, only the *number of personality items* and the VBAs' *creativity* predicted a classification distanced from the mere thing scheme (Table 2, Model 5).

**Mediation Analyses.** Based on the indications of previous studies (e.g., Nass & Lee, 2001; Woods et al., 2007), and because the VBAs' attributes eliminated the effects of the situations' and the users' attributes, we investigated whether mediation effects were involved. We focus on mediated variables that had a significant impact on the classification in Models 1 to 3, verified by the second sample (see below), and on mediating variables which were significantly affecting these. Results are reported in Table A2 (OSF).

*Alexa's* classification increasingly distanced from the pole thing, due to a higher perceived capability of understanding the user compared to the Google Assistant. However, the explained variance is very low, thus the two agents do not differ very much from each other in this respect. Regarding the effect of the *situation*, both ownership and the previous use of the VBAs in the absence of others were mediated by agent attributes. Whereas ownership increased the VBAs' similarity in behaving and thinking and its attributed attentiveness, previous use in the absence of other people heightened all agency and mind attributes of the VBA (except similar feeling), encouraging a classification distanced from the thing scheme. *User attributes* exhibit mediation effects for age and extraversion, both increasing a classification toward the pole thing. Whereas increasing age lowered the attributed attention, number of personality items in general, and the VBAs' creativity, more extraverted users perceived less similarity in feeling and fewer personality items. Effects of matching the respondent's and VBA's personality indicated by previous findings (Nass & Lee, 2001; Woods et al., 2007) could not be confirmed.

Based on the results of the stepwise robust regression and the mediation analysis, the theoretical model was adapted (see Figure A1 in the OSF).

### Validation of the Results by the Staff Sample

In this section, we will focus on the most important commonalities and differences in relation to the student sample. The staff sample confirmed the VBAs' *classification* (Table 1), the amount of *confidence* ($M = 4.45$, $SD = 0.84$) and its relation to the classification, $r = -.48$, $t(429) = -11.38$, $p < .001$. The influences on the classification were partly confirmed (Table A3 in the OSF). Whereas, the impacts of age and extraversion (Model 3), the VBAs' similar behaving (Model 4) and thinking, its creativity, and the interaction of similar thinking and feeling (Model 5) were confirmed, the influences of Alexa (vs. Google Assistant), ownership and previous use in the absence of other people on the classification were not (Model 2). However, the direction of the ownership's and dyad's effect remained and—consistent with the student sample—mediating effects of the VBAs' attributes occurred, although some involved attributes differed (Table A4 in the OSF). The widespread impact of the previous dyadic use on the VBAs' attention, similar thinking and feeling, and creativity was strengthened; the impacts of the users' age through decreased attributed attention to the VBA and of the users' extraversion through decreased perceived similarity in feeling were confirmed.

## Discussion

In this paper we analyzed how people classify their counterpart when they interact with voice-based agents and how attributes of the situation, the user, and the agent influence this classification. By referring to Piaget (1970/1974), we introduced an empirically measurable gradual ontological object–subject classification of VBAs, based on a 100-point scale ranging from thing to person, enabling the identification of the degree to which artificial agents are objects of doubt (Geser, 1989; Reeves & Nass, 1996; Turkle, 1984/2005). Using the VBA as an example, we have demonstrated the potential of this scale, providing the basis for systematic investigations into how people classify various machines and how the classification is affected.

Consistent with previous research (e.g., A. Edwards, 2018; Guzman, 2015), the majority of our participants could not definitely classify VBAs. The object–subject classification ranged between the two schemes thing ("what") and person ("who"), as predicted by Geser (1989) and Gunkel (2020). That is, most people were indeed in doubt. How people compensated for this doubt was analyzed through the concept of equilibration (Piaget, 1970/1974): Whereas some people had assimilated VBAs by objectifying, almost none had subjectified them. However, most people had accommodated their classification by modifying the thing scheme, and some even hybridized it, but still with a bias toward the thing scheme. That is, people reaffirmed their lines between things and persons (Turkle, 1984/2005, p. 34) by embedding VBA in the world of things, partly enriched with aspects of the person scheme. Hence, VBAs are *personified things*.

By understanding that people classify VBAs as personified things, counterintuitive findings (Gambino et al., 2020) can be interpreted more precisely. The thing aspect in the classification *things* and personified *things*, on the one hand, emphasized the VBA's artificiality (see Guzman, 2015), thus encourages the reference "it" (see Purington et al., 2017), or the separation of machines from humans and apes (see A. Edwards, 2018). The person aspect of *personified* things, on the other hand, emphasized the VBA's personhood, causing the simultaneous use of the pronouns "it" and "she" (Guzman, 2015; Purington et al., 2017) and the assignment of social machines to (living) beings (see A. Edwards, 2018). Whereas the aspect of personhood may cause social reactions toward artificial agents (see Appel et al., 2012; Horstmann et al., 2018), their dominant nature as things is likely to be responsible for rather timid, and normatively undesirable social reactions, such as insults or discourtesy (see Cercas Curry & Rieser, 2018; C. Edwards et al., 2019).

However, the classification of personified things was characterized by high *uncertainty*. The further the participants moved away from the established thing or person scheme, the less confident they became about their classification. That is, classifying by assimilation is the easy way, it is the passive (habituated) assignment of an object to an existing scheme. Whereas accommodation—as an active cognitive process of reaffirming the boundaries of the schemes—is fraught with far more doubts. As a result, this classification of personified things is fragile and unstable and may change significantly in time—especially regarding further developments of the agents' abilities and their societal embedding. Further longitudinal research will be required.

How much VBAs were classified as personified things was—in accordance with previous research (e.g., A. Edwards, 2018; Moon et al., 2016; Nass & Brave, 2005)—mainly affected by the perception of the agent's attributes of *agency* (behaves similarly, is attentive) and *mind* (thinks and feels similar, understands, has a vast personality, especially a creative one). However, for the same effect on the classification, it required a much lower degree of mind than agency. Hence, "personified things" expands "social things" (Guzman, 2015) by implying abilities associated with personhood—even if their degree is low (Hubbard, 2011).

However, attributes of the *situation* and the *user* did indirectly alter the classification by affecting the VBAs' assumed agency and mind. In accordance with prior research (A. Edwards et al., 2019; Leite et al., 2013), previous regular interactions positively affected the perception of VBAs as personified things. In contrast to previous studies (e.g., Purington et al., 2017), the main use of VBAs in the *dyad* situation (i.e., in the absence of others) caused a stronger classification toward *personified things*, through increased perceived agency *and* mind. One explanation for this may be that the dyad fosters a more intense experience and lacks the "blueprint" of a human being. As a result, the users can engage more effectively with the VBA but have only themselves for comparison to project subjectivity. The effect found by Purington et al. may be caused by different dynamics. Thus, we encourage to further explore the dynamics of dyadic and multi-person environments regarding the classification.

Corresponding with Piaget (1970/1974), *age* was a crucial factor for subjectification by altering the VBA's agency and mind (e.g., Woods et al., 2007): The younger the participants, the more attentive, creative, and substantial in personality the VBAs were perceived to be. An essential factor in this respect is the continually increasing experience with objects of the environment, which gradually complement the classification and thus extend the options for comparison of included objects in the scheme.

Rather than an effect of matching *personalities* of the user and the VBA (Nass & Lee, 2001), the findings were more ambiguous. Although VBAs were seen as very extraverted across subgroups, introvert people subjectified them more often, through increased perception of similarity in feeling and personality items in general. That is, introverts perceived a richer personality and emotionality in VBAs. Again, this phenomenon can be related to the above-mentioned projection of the users' experienced own subjectivity (Piaget, 1970/1974). As introverts may experience themselves as rich in emotion and personality, even though they are—compared to extroverts—less inclined to express these qualities to others, they likely attribute the same intensity of feelings and personality to the weak expressions of the VBAs.

As with all research, this study has several limitations, of which its hypothetical interaction with its pre-defined questions and answers is one. Additional studies need to validate the VBA's classification in real interactions and with different contents. A second limitation is the paper's focus on VBAs. Comparing various evocative technologies may allow a more distinct classification and may reveal differences in influencing attributes. Third, the classification of an object is the result of a dynamic equilibration process (Piaget, 1970/1974). Hence, we solely examined a snapshot of the process, not the process itself. However, it is unclear whether the classification resulted from initial or multiple equilibration processes. Most of the participants had at least some experience with the VBA, indicating potential prior equilibrations. Further research on the process itself, concerning initial and repeated equilibrations, influences, and related classifications, is needed.

To conclude, human-machine communication in the context of voice-based agents indicates that people communicate with *personified things* defined by a moderate agency and a basic mind. The more VBAs behave, think, feel similar, are perceived to be attentive to, and understand their users, and are at least to some extent creative, the more they are classified as personified things. Although this requires a moderate agency, a rather limited mind is sufficient. Depending on the opportunities for comparison on different levels, the classification is more or less hybrid. Comparisons take place on an individual (age, personality) and situational level (dyad). Age relates to the amount of experience with comparable evocative objects, users' personality relates to the comparison with their own behavior and mind, and the dyadic situation relates to the comparison with another human subject. However, the classification of VBAs as personified things is still fragile and they remain objects of doubt.

## Author Biographies

**Katrin Etzrodt** (MA) is a research assistant and PhD student at the Chair of Science and Technology Communication at the Technical University Dresden. From 2017 to 2020 she was granted a scholarship of the Program for the Promotion of Early-Career Female Scientists of TU Dresden.

https://orcid.org/0000-0001-6515-9985

**Sven Engesser** (PhD) is a Professor of Science and Technology Communication at the Technical University Dresden.

https://orcid.org/0000-0003-1638-7548

## Acknowledgments

The authors would like to thank the reviewers and the editorial team, whose suggestions have considerably improved the paper. Additionally, the authors would like to thank Lisa Weidmüller for her input and feedback at different stages of development. Ms. Etzrodt gratefully acknowledges the support and generosity of the Scholarship Program for the Promotion of Early-Career Female Scientists of TU Dresden, without which the present paper would not have been possible.

The supplemental material can be obtained from the study's Open Science Framework (OSF) repository: https://doi.org/10.17605/OSF.IO/5UNS2

## References

Appel, J., von der Pütten, A., Krämer, N. C., & Gratch, J. (2012). Does humanity matter? Analyzing the importance of social cues and perceived agency of a computer system for the emergence of social reactions during human-computer interaction. *Advances in Human-Computer Interaction*. https://doi.org/10.1155/2012/324694

Attig, C., Wessel, D., & Franke, T. (2017, July 9–14). *Assessing personality differences in human-technology interaction: An overview of key self-report scales to predict successful interaction.* Proceedings of HCI International, Vancouver, BC, Canada. https://doi.org/10.1007/978-3-319-58750-9_3

Biocca, F., Harms, C., & Burgoon, J. (2003). Towards a more robust theory and measure of social presence: Review and suggested criteria. *Presence, 12*, 456–480. https://doi.org/10.1162/105474603322761270

Broadbent, E. (2017). Interactions with robots: The truths we reveal about ourselves. *Annual Review of Psychology*, *68*(1), 627–652. https://doi.org/10.1146/annurev-psych-010416-043958

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Ramirez, Jr., A., Dunbar, N. E., & Miczo, N. (1999). Testing the interactivity model: Communication processes, partner assessments, and the quality of collaborative work. *Journal of Management Information Systems, 13*(6), 33–56. https://doi.org/10.1080/07421222.1999.11518255

Buss, D. M. (2008). Human nature and individual differences. Evolution of human personality. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality. Theory and research* (3rd ed., pp. 29–60). The Guilford Press.

Cercas Curry, A., & Rieser, V. (2018, June). *#MeToo Alexa: How conversational systems respond to sexual harassment*. Proceedings of ACL Workshop on Ethics in Natural Language Processing. https://doi.org/10.18653/v1/W18-0802

Dimock, M. (2019). Defining generations: Where Millennials end and Generation Z begins. Pew Research Center. http://tony-silva.com/eslefl/miscstudent/downloadpagearticles/defgenerations-pew.pdf

Edwards, A. (2018). Animals, humans, and machines: Interactive implications of ontological classification. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 29–49). Peter Lang.

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior*, *90*, 308–314. https://doi.org/10.1016/j.chb.2018.08.042

Edwards, A., Westerman, D., Edwards, C., & Spence, P. R. (2020). Communication is . . . transhuman. In A. Tyma & A. Edwards (Eds.), *Communication is … Perspectives on theory* (pp. 49–66). Cognella Academic Publishing.

Edwards, C., Edwards, A., Stoll, B., Lin, X., & Massey, N. (2019). Evaluations of an artificial intelligence instructor's voice: Social identity theory in human-robot interactions. *Computers in Human Behavior*, *90*, 357–362. https://doi.org/10.1016/j.chb.2018.08.027

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Etzrodt, K. (2021). The ontological classification of conversational agents. An adaptation of Piaget's equilibration theory. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, E. Luger, M. Goodwin, & P. B. Brandtzaeg (Eds.), *CONVERSATIONS 2020—4th International Workshop on Chatbot Research*. Lecture Notes in Computer Science. Springer.

Franke, T., Attig, C., & Wessel, D. (2018). A personal resource for technology interaction: Development and validation of the affinity for technology interaction (ATI) scale. *International Journal of Human–Computer Interaction, 35*(6), 456–467. https://doi.org/10.1080/10447318.2018.1456150

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, 1, 71–85. https://doi.org/10.30658/hmc.1.5

Garcia, D. M. P., Lopez, S. S., & Donis, H. (2018, July). *Voice activated virtual assistants personality perceptions and desires: Comparing personality evaluation frameworks*. Proceedings of International BCS Human Computer Interaction Conference, Swindon United Kingdom. https://doi.org/10.14236/ewic/HCI2018.40

Geser, H. (1989). Der PC als Interaktionspartner [The PC as interaction partner]. *Zeitschrift für Soziologie*, *18*(3), 230–243. https://doi.org/10.1515/zfsoz-1989-0305

Guadagno, R. E., Blascovich, J., Bailenson, J. N., & McCall, C. (2007). Virtual humans and persuasion: The effects of agency and behavioral realism. *Media Psychology*, *10*(1),1–22.

Gunkel, D. J. (2020). *An introduction to communication and artificial intelligence*. Wiley.

Guzman, A. L. (2015). *Imagining the voice in the machine: The ontology of digital social agents* [PhD Thesis]. University of Illinois at Chicago.

Guzman, A. L. (2018). What is human-machine communication, anyway? In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang. https://doi.org/10.3726/b14399

Guzman, A. L. (2019). Voices in and of the machine: Source orientation toward mobile virtual assistants. *Computers in Human Behavior*, *90*, 343–350. https://doi.org/10.1016/j.chb.2018.08.009

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for Human-Machine Communication. *Human-Machine Communication, 1*, 37–54. https://doi.org/10.30658/hmc.1.3

Higgins, E. T., & Scholer, A. A. (2008). When is personality revealed? A motivated cognition approach. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of Personality. Theory and Research* (3rd ed., pp. 128–207). The Guilford Press.

Horstmann, A. C., Bock, N., Linhuber, E., Szczuka, J. M., Straßmann, C., & Krämer, N. C. (2018). Do a robot's social skills and its objection discourage interactants from switching the robot off? *PLoS ONE*, *13*(7): e0201581. https://doi.org/10.1371/journal.pone.0201581

Hubbard, F. P. (2011). "Do androids dream?": Personhood and intelligent artifacts. *Temple Law Review, 83*, 405–474. https://ssrn.com/abstract=1725983

Knorr-Cetina, K. D. (2001). Objectual practice. In T. R. Schatzki (Ed.), *The practice turn in contemporary theory* (pp. 175–188). Routledge, London.

Krummheuer, A. (2015). Technical agency in practice: The enactment of artefacts as conversation partners, actants and opponents. *PsychNology Journal*, *13*, 179–202. http://www.psychnology.org/File/PNJ13%282-3%29/PSYCHNOLOGY_JOURNAL_13_2_KRUMMHEUER.pdf

Lang, D. S. (2008). *Soziale Kompetenz und Persönlichkeit. Zusammenhänge zwischen sozialer Kompetenz und den Big Five der Persönlichkeit bei jungen Erwachsenen* [Social competence and personality. Correlations between social competence and the Big Five of personality in young adults] [PhD Thesis]. Universität Koblenz-Landau. https://kola.opus.hbz-nrw.de/opus45-kola/frontdoor/deliver/index/docId/264/file/SozialeKompetenzundPersoenlichkeit.pdf

Latour, B. (1995). *Wir sind nie modern gewesen. Versuch einer symmetrischen Anthropologie* [We have never been modern]. Suhrkamp. (Original work published 1991.)

Latour, B. (2005). *Reassembling the social. An introduction to Actor-Network-Theory (ANT)*. Oxford.

Lee, E. J. (2008). Flattery may get computers somewhere, sometimes: The moderating role of output modality, computer gender, and user gender. *International Journal of Human-Computer Studies*, *66*(11), 789–800. https://doi.org/10.1016/j.ijhcs.2008.07.009

Lee, E. J., Nass, C., & Brave, S. (2000, April). *Can computer-generated speech have gender? An experimental test of gender stereotype.* Proceedings of CHI on Human Factors in Computing Systems, the Hague the Netherlands. https://doi.org/10.1145/633292.633461

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

Lee, K. M., & Nass, C. (2005). Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology*, *7*(1), 31–45. https://doi.org/10.1207/S1532785XMEP0701_2

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: A survey. *International Journal of Social Robotics 5*, 291–308. https://doi.org/10.1007/s12369-013-0178-y

Lombard, M., Lee, S., Sun, W., Xu, K., & Yang, H. (2017). Presence theory. In P. Rössler, C. A. Hoffner, & L. van Zoonen (Eds.), *The international encyclopedia of media effects* (pp. 1–13). John Wiley & Sons Inc. https://doi.org/10.1002/9781118783764.wbieme0087

Lopatovska, I., & Williams, H. (2018, March 11–15). *Personification of the Amazon Alexa: BFF or a mindless companion?* Proceedings of the ACM CHIIR Conference, New Brunswick. https://doi.org/10.1145/3176349.3176868

Moon, Y., Kim, K. J., & Shin, D.-H. (2016, July 17–22). *Voices of the Internet of Things: An exploration of multiple voice effects in smart homes*. In P. M. Norbert Streitz (Ed.), Proceedings of the DAPI Conference held as part of HCI International, Toronto, Canada. https://doi.org/10.1007/978-3-319-39862-4

Nass, C., & Brave, S. (2005). *Wired for speech—How voice activates and advances the human-computer relationship*. MIT Press.

Nass, C., & Lee, K.-M. (2001). Does computer-generated speech manifest personality? Experimental test of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied, 7*(3), https://doi.org/10.1037/1076-898X.7.3.171

Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56*(1), 88–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Steuer, J., & Tauber, E. R. (1994). *Computers are social actors.* Proceedings of CHI Conference on Human Factors in Computing Systems, Boston Massachusetts USA. https://doi.org/10.1145/191666.191703

Newman, N. (2018). *Digital news project. The future of the voice and the implications of news*. Reuters Institute for the Study of Journalism & University of Oxford. https://reutersinstitute.politics.ox.ac.uk/our-research/future-voice-and-implications-news

Piaget, J. (1974). *Abriß der genetischen Epistemologie* [The principles of genetic epistemology] (F. Kubli, Trans.). Walter-Verlag AG Olten. (Original work published 1970.)

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017, May). "Alexa is my new BFF": Social roles, user satisfaction, and personification of the Amazon Echo. *CHI'17 Extended Abstracts*. https://doi.org/10.1145/3027063.3053246

Ramirez Jr., A. (2009). The effect of interactivity on initial interactions: The influence of information seeking role on computer-mediated interaction. *Western Journal of Communication, 73*(3), 300–325. https://doi.org/10.1080/10570310903082040

Rammstedt, B., Kemper, C. J., Klein, M. C., Beierlein, C., & Kovaleva, A. (2013). A short scale for assessing the big five dimensions of personality: 10 item big five inventory (BFI-10). *Methoden, Daten, Analysen, 7*(2), 233–249. https://doi.org/10.12758/mda.2013.013

Ray, G. B. (1986). Vocally cued personality prototypes: An implicit personality theory approach. *Communication Monographs*, *53*(3), 266–276. https://doi.org/10.1080/03637758609376141

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

Reichertz, J. (2014). Von Menschen und Dingen. Wer handelt hier eigentlich [Of people and things. Who is actually acting]? In A. Poferl & N. Schröer, (Eds.), *Wer oder was handelt? Zum Subjektverständnis der hermeneutischen Wissenssoziologie*, 95–120. Springer Fachmedien. https://doi.org/10.1007/978-3-658-02521-2

Roßler, G. (2008). Kleine Galerie neuer Dingbegriffe: Hybriden, Quasi-Objekte, Grenzobjekte, epistemische Dinge [Small gallery of new terms: hybrids, quasi-objects, boundary objects, epistemic things]. In G. Knef, M. Schröer, & E. Schüttpelz (Eds.), *Bruno Latours Kollektive* (pp. 76–107). Suhrkamp.

Schütz, A. (1974). *Der sinnhafte Aufbau der sozialen Welt: Eine Einleitung in die verstehende Soziologie* [The phenomenology of the social world: An introduction into verstehende sociology]. Suhrkamp, Frankfurt.

Short, J., Williams, E., & Christie, B. (1976). *The social psychology of telecommunications*. Wiley.

Simmel, G. (1908). *Soziologie Untersuchungen über die Formen der Vergesellschaftung* [Inquiries into the construction of social forms]. Duncker & Humbolt.

Singh, A. (2014). Challenges and issues of Generation Z. In *Journal of Business and Management, 16*(7), 59–63. https://pdfs.semanticscholar.org/b109/73a5c6d11f37542adc34455bb0c7fbcbbb9d.pdf

Suchman, L. (2011). Subject objects. *Feminist Theory, 12*(2), 119–145. https://doi.org/10.1177/1464700111404205

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460. https://www.jstor.org/stable/2251299

Turkle, S. (2005). *The second self (Twentieth anniversary edition)*. The MIT Press. (First edition published by Simon & Schuster, Inc., New York, 1984). https://doi.org/10.7551/mitpress/6115.001.0001

Weber, M. (1922). *Wirtschaft und Gesellschaft* [Economy and society]. Mohr, Tübingen.

Woods, S., Dautenhahn, K., Kaouri, C., te Boekhorst, R., & Koay, K. L. (2007). Is this robot like me? Links between human and robot personality traits. *Interaction Studies, 8*(2), 281–305. https://doi.org/10.1075/is.8.2.06woo

HMC

HUMAN-MACHINE

COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.4

# Forms and Frames: Mind, Morality, and Trust in Robots Across Prototypical Interactions

**Jaime Banks[1], Kevin Koban[2], and Philippe de V. Chauveau[1]**

1 College of Media & Communication, Texas Tech University, Lubbock, Texas, USA
2 Department of Communication, University of Vienna, Währinger Straße 29, 1090 Vienna

## Abstract

People often engage human-interaction schemas in human-robot interactions, so notions of prototypicality are useful in examining how interactions' formal features shape perceptions of social robots. We argue for a typology of three higher-order interaction forms (social, task, play) comprising identifiable-but-variable patterns in agents, content, structures, outcomes, context, norms. From that ground, we examined whether participants' judgments about a social robot (mind, morality, and trust perceptions) differed across prototypical interactions. Findings indicate interaction forms somewhat influence trust but not mind or morality evaluations. However, how participants *perceived* interactions (independent of form) were more impactful. In particular, perceived task interactions fostered functional trust, while perceived play interactions fostered moral trust and attitude shift over time. Hence, prototypicality in interactions should not consider formal properties alone but must also consider how people perceive interactions according to prototypical frames.

**Keywords:** social cognition, trust, schema, framing, prototypicality, playfulness, temporality

## Introduction

News and popular media depict potentials for social robots (embodied technologies that simulate contextually social behaviors; Duffy, 2003) to interact with humans as do other

This material is based upon work supported by the Air Force Office of Scientific Research under award number FA9550-19-1-0006.

**CONTACT** Jaime Banks • College of Media & Communication • Texas Tech University • Box 43082 • Lubbock, TX 79409 • j.banks@ttu.edu

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

humans. Those interactions comprise distinct orientations, dynamics, and outcomes that may fundamentally influence humans' understandings of robots. In particular, since specific interactions may be understood as prototypical of common interaction types, people may draw on scripts or schema inherent to those forms when evaluating social robots. To better understand how interaction types may shape perceptions of robots, we draw on human communication literature to argue for an organizing framework of higher-order *interaction forms* and their prototypical features. We employ this framework to isolate the particular features of relational norms and aims and to empirically examine their influence on social perceptions of a robot.

Thus, the goals of this research are twofold: to (a) conceptualize interactions as prototypical of interaction forms and (b) investigate those forms' impact on robot perceptions. To these ends, we explicate the notion of prototypicality and how it emerges in interactions, and then argue for three higher-order interaction forms (social, task, play) comprising distinct sets of commonly co-occurring features. On that foundation, we present hypothesis and formulate questions regarding interaction forms' impacts on key social(-cognitive) evaluations of robots. After reporting experiment results, findings are discussed in relation to prior research on robot schemas, interactants' individual frames for interaction forms, and temporality of social perceptions. Findings indicate that mind perception and moral judgment did not vary across interaction forms; trust differed mildly across *manipulated* interaction forms, but more strongly across *perceived* interaction forms, suggesting prototypicality is a matter of both formal and perceived properties.

## Literature Review

Human communication scholarship indicates that individual-, group-, and societal-level differences can influence how meaning is made in communication processes. In order to find superordinate patterns that permit generalizable ideas about how communication unfolds, scholars often rely on the notion of *prototypicality.* Prototypicality is the extent to which an exemplar is a good representation of its category (Rosch, 1973). This tautological definition belies the construct's dynamic nature as prototypicality relies on frequencies of exemplar encounters and perceptions of each exemplar's representativeness (Nedungadi & Hutchinson, 1985). In other words, the more people encounter an exemplar and perceive it as representative of a category, the more likely that exemplar will emerge as prototypical of the category.

It may be tempting to consider prototypical exemplars, then, as comprising clear sets of features. For instance, one might say an interaction is "professional" if it has necessary features of formality, hierarchical relations, and workplace context. However, prototypicality in communication is better considered as having fuzzy-set criteria: an exemplar's representativeness of a category varies with the *degree to which* it adheres to a collection of sufficient-but-not-necessary qualities (see Ragin, 2000). Thus, a "professional" interaction is instead one that is *sufficiently* formal, hierarchical, and/or workplace-situated. Non-prototypicality would equate to "0" and perfect prototypicality would equate to "1," with grades between. Although prototypicality is principally examined in individuals' perceptions of other people, the notion is a useful conceptual starting point for forms of interaction.

## Prototypicality in Interaction

Prototypicality emerges through systematically shared features (Nedungadi & Hutchinson, 1985) that cumulatively guide people's shared understanding of the world. Accordingly, specific interaction forms (i.e., higher-order categories) may be identified by considering features that commonly co-occur when agents interact in exemplar interactions. Identifying such features is perhaps best grounded in a long-standing definition of communication: Who says What to Whom in which Channel with what Effect (Lasswell, 1948). Who(m) refers to the *agents* involved, What is the *content* of the messages, Channel includes the *structural features* of an interaction, and Effect is the *outcome*. As all communication is embedded in social situations governed by contextual factors and social rules (Goffman, 1964), we add to the Lasswellian formula *context* and *norms*. Higher-order interaction forms can be identified, then, by common agents, content, structures, outcomes, context, and norms. Specific interactions are more or less prototypical of those forms as they are more or less adherent to the common features.

These six clusters of interaction features set the scope for further explicating systematic feature co-occurrence (see Figure 1). There is systematicity in *agents* in terms of their "kind," where kind is discernible in demographics (e.g., gender expression), social category (e.g., blue-collar workers), or ontological category (e.g., robots; Guzman, 2020). Systematicity in *content* includes feature patterns in the subject matter and topic of interactions (Taboada & Wiesemann, 2010). Systematicity in *structures* comprise the composition of messages (How does an interaction unfold?) and channels (How is an interaction arranged?). Systematicity in *outcomes* is broadly construed as the individually or jointly (un)expected or (un)desired results of an interaction, such as goals (Homans, 1961) and effects (Altman & Taylor, 1973). Systematicity in *context* consists of the spatial and social environments (Bronfenbrenner, 1977). Finally, systematicity in *norms* are implicit or explicit social rules for how individuals have or ought to behave in an acceptable fashion (Lapinski & Rimal, 2005).

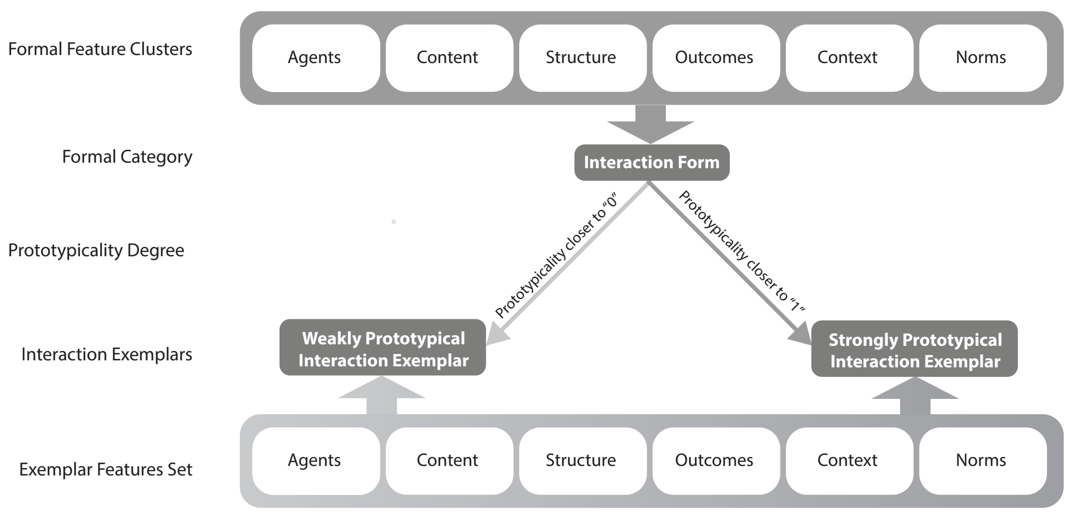



**FIGURE 1 Formal and Operational Layers of a Model for Interaction-Form Prototypicality**

## Higher-Order Interaction Forms: Social, Task, Play

We have argued that six clusters of features have subsets that commonly co-occur across interaction exemplars. These co-occurrence patterns give rise to higher-order interaction forms—akin to how others have proposed operational user-agent interaction "cases" (e.g., deception, coercion) based on technical parameters choices, learning; Burr et al., 2018). We build on the notion of interaction-form prototypicality to argue for *higher-order* forms across which those cases (and others) may manifest. Prompted by exploratory work suggesting differential social-cognitive effects across social, task, and playful interactions (Banks, 2017), literature regarding those three interactions suggests variably co-occurring features and so candidacy as higher-order interaction forms.

**Social interaction.** Social interactions, broadly, are processes by which agents co-act (Mele, 2017). Under this definition, *most* interaction exemplars may be considered social. However, we refer specifically to interactions in which sociality is an end in itself. Co-occurring features manifest in relation to *agents* (widely varied, but generally of comparable power; Dahl, 1957), *content* (self- and other-relevant subjects such as life stories and opinions; see Mehl & Pennebaker, 2003), *structures* (reciprocity, turn-taking, increasing self-disclosures; Rubin & Shenker, 1978), *outcomes* (feelings of relatedness, closeness, familiarity, kinship; Aron et al., 1992), *context* (often third places or personal spaces during leisure time; Oldenburg, 2007), and *norms* (manners, attention, role-taking; Kurth, 1970).

**Task interaction.** Task interactions are those constellated around exertion of mental or physical resources, primarily in service of an interested entity (Burke, 1971). That interested entity may be oneself (a person training for a marathon with an AI coach), another (an electrician working for a homeowner) or an external entity (one performing cooperative labor for an employing company). Co-occurring features manifest in relation to *agents* (actors with complementary skill sets and authorities, generally motivated toward mutual benefit; see Cummings & Kiesler, 2008), *content* (goal- and process-relevant subjects; see Pickering & Garrod, 2006), *structures* (standardized communication protocols, efficient channels, established timelines for repeated contact; Bolton, 2015), *outcomes* (achievement of the aim, feelings of competence, rewards for performance; Weiss & Kahn, 1960), *context* (co-located or mediated work spaces, hierarchical social structures; Poole, 1978), and *norms* (shared or tiered responsibility, distribution of roles, collaboration; Teh et al., 2012).

**Play interaction.** Playful interactions are autotelic communications situated outside of (albeit in relation to) everyday work or social life (cf. Huizinga, 1949). Although the notion of play may invoke strong associations with games (i.e., [semi-]structured challenge systems with negotiable outcomes; Juul, 2005), playful interactions also include other autotelic communicative activities (e.g., those in leisure, work, education, or culture; Frissen et al., 2015) situated within a collectively pretended, superimposed system of rules (Frasca, 2007) and characterized by a mindset of gregariousness, frivolity, and reduced inhibition (see Barnett, 2007). Co-occurring features manifest in relation to *agents* (familiars with shared justification for the activity; Deterding, 2018), *content* (subjects related to the activity itself or communication that is silly, whimsical, funny, or teasing; Mäyrä, 2012), *structures* (game- or participant-defined frameworks for collaboration, competition, or creation; Boyan &

Banks, 2018), *outcomes* (hedonic and eudaimonic gratifications; Rogers et al., 2017), *context* (private spaces or formalized public play spaces; see Foucault, 1986), and *norms* (e.g., accordance with or free from rules or boundaries, suspension of disbelief; Burke, 1971).

## Interaction Forms and Social Perceptions of Robots

Social, task, and play interaction forms afford a tentative typology, and specific interaction exemplars (i.e., given conversations) variably fit into the higher-order categories based on adherence to some number of sufficient-but-not-necessary features. In this way, the forms are concrete enough to discern across more or less prototypical exemplars but still flexible enough to shift with emergent communication activities. Some exemplars are clearly and largely prototypical of an interaction form (most/all common features, closer to full prototypicality of "1") while others have some degree of fit depending on the number of representative features (some features, prototypicality between 0 and 1). For instance, considering playful interactions, a robot playing soccer with a human in a park fits many aforementioned features for prototypical play interactions; however, the same pair playing in a grocery store has lesser fit (as the context is not prototypical). This concrete-yet-flexible typology is useful for systematically exploring the influence of interactions—as complex feature-sets—on relational outcomes.

Prototypicality informs individuals' schemas and frames that guide their expectations for an interaction. Schemas are "superordinate knowledge structures that reflect abstracted commonalities across multiple experiences," shaping how people interpret new experiences (Gilboa & Marlatte, 2017, p. 618). Individually held frames are schema-derived filters that organize immediate experiences (Goffman, 1974) and serve as interpretive lenses (Scheufele, 2000). In other words, sets of features held together in each higher-order interaction form are—through experience-derived schemas and frames—familiar and recognizable, guiding how people approach novel-but-familiar interactions.

Prototypical social, task, and play interactions foster evolving schemas that include descriptive and procedural scripts as each is encountered in everyday life. Scripts are abstract representations of stereotypical event-sequences, accessed in response to relevant situational cues (Abelson, 1981): sets of ideas about what one and others should do in a situation. People engage in similar social-cognitive processes (e.g., mind perception; Banks, 2020b) and interaction scripts (Edwards et al., 2016) when interacting with robots, though it is yet unclear whether the *form* of an interaction may influence these dynamics. In considering those potential effects, it is useful to first explore impacts on fundamental social judgments in relation to prototypical norms and outcomes.

**Mind perception.** People's perception of another agent as having a mind unfolds through an inferential mechanism (Theory of Mind; Premack & Woodruff, 1978) by which they overtly or indirectly surmise the agent's mental states from discernible cues. Three perceived mental capacities (Malle, 2019) are (a) reality-interaction capacity (mindful communication and action), (b) affective capacity (physiologically/emotionally experiencing emotional states), and (c) social-moral capacity (simulation of others' minds and evaluating right/wrong). Social action often automatically promotes inferences of mindedness and since humans generally mentalize robots as they do humans (Banks, 2020a), making it likely that all three interaction forms will promote some level of all three capacities, albeit with differential patterns. Because social interactions are grounded in self-relevant exchanges via *norms* for self-disclosure that engender relational intimacy *outcomes* (Altman & Taylor, 1973),

> **(H1a)** social interaction will promote highest perceived affective capacity. Task interactions feature shared responsibility and cooperation *norms* (Wageman & Baker, 1997) leading to incremental progress and effective task achievement *outcomes* (Walliser et al., 2019), such that
>
> **(H1b)** task interaction will promote strongest reality-interactive mentalizing. With *norms* for fair play (cf. Consalvo, 2009a) and *outcomes* of coordination-driven achievement and positive affect (Oliver et al., 2015),
>
> **(H1c)** playful interactions will promote highest perceived social-moral capacities. Because this indirect mental inferencing operates independently of more conscious judgments, it is also useful to consider:
>
> **(RQ1)** Does explicit mind ascription differ across interaction forms?

**Moral evaluation.** Perception of entities' moral agency is key to a range of social cognitions (Moll et al., 2008) that may influence human-machine communication. Extrapolating from scholarship on media representations of morality, people show greater affinity for personas that express valued moral-character qualities (Tamborini et al., 2013) and desire punishment for those that deviate (Zillmann & Cantor, 1976). Such moral evaluations may vary by interaction form, as a form's prototypical features make salient the behaviors considered *normatively* "good" in an interaction and the types of *outcomes* that excuse "bad" behaviors. For instance, robots telling lies to humans in social interactions may be seen as immoral due to expectations for authenticity but found acceptable or desirable in play interactions where lies are prototypical (e.g., "poker faces").

In addition to discrete moral-character evaluations, explicit moral status ascription (overt decision on robots' abilities to be or do good/bad) may vary across interaction forms. An interaction form's prototypical *norms* and *outcomes* may give the human partner greater license for acknowledging moral status. However, because moral evaluations depend heavily on the *content* of messages conveyed, behaviors performed, and people's emotional responses to it (Avramova & Inbar, 2013), neutral content (as in the present study's stimuli) could also prevent variation. Given these conflicting potentials, we ask:

> **(RQ2)** (How) does robot (a) character evaluation and (b) moral-status ascription differ across interaction forms?

**Trust(worthiness).** Trust is a multifaceted construct understood as a weak induction (somewhere between full knowing and ignorance) that provokes surrender in the face of uncertainty (Simmel, 1908/1950, 1900/1990). Within human-machine communication research it is defined as feelings of faith and reliance that a robot will not exploit one's vulnerability, and it is understood as key to social acceptance of robots (see Ullman & Malle, 2018) as it may play out in socioemotional distancing and intentions to engage them. Although extant literature suggests playfulness may be associated with positive engagement (Moon & Kim, 2001), a robot's performance may affect explicit ascription of trust (van den Brule et al., 2014), and linguistics may influence social distance (Kim et al., 2013). However, it is yet unclear whether higher-order interaction forms may engender differences in trust. Therefore, we ask:

> (How) do **(RQ3a)** explicit trust ascription and **(RQ3b)** trust-related intentions (i.e., future engagement and preferred social distance) vary across interaction forms?

In social judgments of robots, trust as a feeling-state is different from trustworthiness as the perception of another actor's abilities or character (Colquitt et al., 2007). Robot trustworthiness is multidimensional (Ullman & Malle, 2019), comprising considerations of capacity (whether a robot is functionally capable and operationally reliable) and morality (whether it is inherently ethical and sincere). We predict that both types of trustworthiness will vary by interaction form. As social interactions rely on *norms* of authentic self-disclosure and *outcomes* of relational closeness (Altman & Taylor, 1973) and play interactions rely on *norms* of rule-following and fairness (Consalvo, 2009a), we hypothesize that

> **(H2a)** social and play interactions will promote highest levels of moral trustworthiness. Regarding capacity trustworthiness, competence is known to precede trust in another's capabilities (Sherwood & DePaolo, 2005) and capacity trustworthiness manifests partly through cooperative decision-making (McCabe et al., 2001), which is common to both task and play interactions. Following, it is likely that in comparison to social interactions,

> **(H2b)** task and play interactions will promote highest capacity trustworthiness perceptions.

## Method

The study implemented three exemplars prototypical of the three higher-order interaction forms (social, task, play) as a between-subjects factor. Participants were randomly assigned to one of these conditions, resulting in $n = 40$ participants in the social condition, $n = 38$ in the task condition, and $n = 39$ in the play condition. All study materials are available in online supplements: https://osf.io/n87bg/.

### Participants

A convenience sample of students was invited to participate in a study on "interacting with robots in different situations" for course credit and US$5 compensation. After excluding three participants due to technical issues, impaired vision, and missing survey responses, the final sample comprised 117 participants aged $M = 20.38$ years ($SD = 3.45$, range: 18–40). In total, 43 participants identified as male and 74 as female. Ethnicities included Caucasian

($n$ = 75), Hispanic ($n$ = 19), African ($n$ = 9), Asian ($n$ = 6), native American ($n$ = 1), and mixed ethnicities ($n$ = 7).

## Procedure

The study's procedure consisted of three stages. First, participants completed a pre-session online survey capturing covariates: robot experience, existing robot and technology attitudes, and demographics.

Second, they participated in an in-person lab session to complete the robot-interaction procedure (see supplements for lab configuration). After giving informed consent, the experimenter guided participants into another room and briefly introduced them to the humanoid robot "Ray" (RoboThespian with "Pris" face and "Heather" American English voice, addressed with female pronouns); they were given the chance to introduce themselves. The experimenter briefly left the room as they "got to know each other" to minimize immediate novelty effects; Ray asked participants about academic majors, current studies, and hobbies, followed by a short description of Ray's alleged leisure activities ("reading the news on the internet"). The experimenter then stepped back into the room and provided instructions for the assigned interaction that would unfold (sans experimenter) over the next 5 minutes.

Finally, after those 5 minutes, the experimenter returned and led the participant back to the first room to complete a post-interaction survey out of sight of both robot and experimenter. That survey included items capturing initial impression of Ray and of the interaction, and scales for perceptions of mind, morality, and trust. Participants were thanked, compensated, and dismissed.

### Stimulus Interactions

The robot was controlled using Wizard of Oz methodology: although we told participants that Ray could "do a lot of things, like sensing its environment, listening and responding in conversations," the robot was fully controlled from a hidden lab room. One of three student assistants controlled Ray using 36 pre-defined interaction scripts consisting of verbal and non-verbal responses—including affirmatives (e.g., "Oh, very good."), negatives (e.g., "No, I don't think so."), and ambiguous statements (e.g., "Interesting."), as well as several condition-specific responses. A cover story rationalized limitations when participants wanted to talk with Ray beyond the predetermined script, explaining that Ray was still "learning how to speak to humans well," that some "speech may be a little limited," and they should not "be surprised if she doesn't give the perfect answer to what you say." The controllers were trained for 6 hours and ran tests with naive training participants. Form features associated with agents, structures, content, and context were controlled while prototypical features for norms and outcomes were manipulated (Table 1; see supplements for complete scripts and stimuli).

**Social interaction.** In the social condition, participants engaged in "friendly conversation" with Ray about everyday triangle-shaped things (a morally neutral topic). Instructions emphasized self-disclosures through talk about everyday experiences. Both interlocutors

**TABLE 1 Interaction Form Features Manipulated and Held Constant Across Conditions**

| | Social | Task | Play |
|---|---|---|---|
| **Agents** | Human students interacting with a female-cued robot who speaks, gestures, and indicates agency on-screen | | |
| **Content** | "What things in your life are shaped like triangles?" | "identifying hidden triangle shapes in complex photographs" | "control a triangle and . . . get the triangle into its home" |
| **Structure** | Turn-taking, self-disclosure, vocal messaging, exhibition of action on-screen, escalating complexity/difficulty | | |
| **Outcomes** | "get to know each other" | "accuracy in identifying triangles" | "having fun" |
| **Context** | Sterile lab setting (across a table from Ray, with wall-mounted screen and keyboard/gamepad controls), with the interaction framed in the service of research | | |
| **Norms** | "have a conversation together," "socialize," "sharing your experiences" | "complete a task together," "accuracy," specific instructions | "play a game together," beating levels, challenge and achievement signals |

were instructed to use a keyboard to type out topics discussed (typing was shown via Microsoft Word on a wall-mounted screen) to visibly display Ray's agency in parallel with on-screen activities in other conditions. Ray's condition-specific speech included requests to name triangles, think about a favorite triangle, explain an unfamiliar triangle, and speculate about triangles' importance.

**Task interaction.** In the task condition, participants "completed a task together" with Ray: cooperating on a CAPTCHA-like task to find hidden triangles in photographs. Instructions emphasized coordination and accuracy as goal outcomes. Ray's actions were shown on-screen by a moving cursor and clicking identified triangles. Ray's condition-specific speech included task-relevant suggestions, action narrations, and requests for help.

**Play interaction.** In the play condition, participants were asked to "play a game" with Ray. The game was *DeRu* (InkKit, 2018): a puzzle game in which players cooperate to navigate triangles through obstacles to respective goal states. Instructions outlined game rules, goals, and emphasized turn-taking and having fun. Ray's actions were shown on-screen through the control of the triangle game piece, and condition-specific speech included movement suggestions and narratives of actions.

## Measures

All within-scale items responses were presented in a randomized order on 7-point Likert scales and averaged by subscale for further statistical analyses, unless otherwise indicated.

**Mind perception.** Participants were asked to complete the updated 20-item version of the multidimensional mental capacity scale (Malle, 2019) comprising three dimensions: reality-interaction capacity (α = .86), affective capacity (α = .86), and social/moral capacity (α = .86). Participants also rated the robot on the four-item dependency (i.e., mindlessness) subscale of the perceived moral agency scale (α = .53; Banks, 2019) and were asked to explicitly indicate whether or not "Ray has a mind" (no/yes). Due to low internal reliability, the dependency subscale was excluded from further analyses.

**Moral evaluation.** The perceived moral agency scale's six-item moral-capacity subscale (α = .89; Banks, 2019) was employed, complemented by the extended character morality questionnaire (Grizzard et al., 2020) consisting of four-item moral-foundation subscales: care (α = .85), fairness (α = .77), loyalty (α = .77), authority (α = .88), purity (α = .69). The purity subscale was adapted (in cooperation with the scale developer), replacing one item about smoking with one about a computer virus. Two items were dropped from analysis ("healthy lifestyle" from the purity set, "loyalty to friends" from the loyalty set) for low intra-dimension correlations ($r \leq -.02$). Explicit moral status was measured by asking whether or not "Ray can be moral or immoral" (no/yes).

**Trust(worthiness).** Participants' trust in Ray was operationalized as both trustworthiness and trust-related intentions. Trustworthiness was captured via the multidimensional measure of trust (Ullman & Malle, 2019) with two dimensions: capacity trustworthiness (α = .85) and moral trustworthiness (α = .91). Participants also responded to a single dichotomous measure, asking "whether or not they trust Ray" (no/yes). Trust-related intentions were captured via two measures concerning comfort with future interactions. First, physical, relational, and conversational indicators of social distance were employed using three Guttman-scale items (Banks & Edwards, 2019), range 0–6 with lower scores indicating more distance. Second, participants indicated degrees of willingness to meet with Ray again in three different scenarios: "to have a friendly chat," "to work on a difficult problem," or "to play a game."

**Manipulation check.** Although stimulus interactions were carefully constructed according to prototypicality criteria, interactions within laboratory settings may be seen as a task in itself, irrespective of its features. Following, a single-item manipulation check was used to capture perceptions of the interaction as most similar to "having a conversation," "completing a task," or "playing a game." Participants also were asked to describe their experience in an open-ended format.

**Control measures.** Participants documented attitudes toward robots via the Godspeed questionnaire likability subscale (α = .90; Bartneck et al., 2009) and prior experience with robots via a single Likert-style item. General attitudes regarding novel technologies were captured using technophobia (α = .77) and technophilia scales (α = .88; Martínez-Córcoles et al., 2017). Specific to the stimuli, Ray's perceived humanlikeness and amiability were captured using the Godspeed anthropomorphism (α = .83) and likeability subscales (α = .88).

## Results

Analyses draw from a sample of $N = 117$; however, due to survey randomization error, the first 44 cases collected were only shown two of three dependent variable sets, so they suffer from data missing at random. See supplements for specific missing cases, descriptive statistics, and zero-order correlations. In checking effectiveness of the manipulation, participants perceived all conditions as most similar to completing a task (see supplements). However, Chi-Squared testing across conditions demonstrated that no misaligned perceptions were significant (*p*s ≥ .150). Based on these discrepancies between assigned and perceived interaction form, analyses evaluated both manipulated and perceived interaction forms as between-subjects factors.

### Hypothesis Testing

In multivariate and univariate tests (as appropriate, given [non]correlations between dependent variables), when significance was approached (*p* range .051–.100), moderate or larger effect sizes (Cramér's $V \geq .21$ [given $df = 2$] and univariate part. $\eta^2 \geq .06$) were interpreted, given smaller sample sizes inappropriately punish *p* values (Bowman, 2017).

**Mind perception (H1a–c, RQ1).** MANCOVA compared mental capacities across dimensions; participants' general technology and robot attitudes and Ray's perceived humanlikeness and amiability were covariates. There was no significant multivariate effect for mental capacities, Wilks' $\lambda = .97$, $F(6,192) = 0.46$, $p = .837$. Therefore, hypotheses H1a, H1b, and H1c were not supported. A Chi-Square test compared explicit mind ascription across conditions, with no significant differences, $\chi^2(2) = 3.08$, $p = .215$, $V = .171$. Answering RQ1, then, explicit mind ascription did not significantly differ across interaction forms.

**Moral evaluation (RQ2).** A MANCOVA compared moral character and ANCOVA compared moral status across conditions, with the same covariates as above. There was neither a significant multivariate effect, $\lambda = .86$, $F(10,188) = 1.45$, $p = .161$, nor a significant main effect, $F(2,98) = 1.29$, $p = .279$, part. $\eta^2 = .026$. Concerning explicit moral-status ascription across conditions, analysis again revealed no significant differences, $\chi^2(2) = 0.30$, $p = .860$, $V = .053$. Answering RQ2, neither implicit nor explicit moral evaluations varied across manipulated interaction forms.

**Trust(worthiness) (H2a–b, RQ3a–b).** MANCOVA (with covariates as above) compared trustworthiness across conditions. Although there was a significant multivariate effect, $\lambda = .87$, $F(4,194) = 3.44$, $p = .010$, subsequent univariate testing revealed no significant differences (capacity trustworthiness: $F(2,98) = 2.82$, $p = .064$, part. $\eta^2 = .054$; moral trustworthiness: $F(2,98) = 2.94$, $p = .058$, part. $\eta^2 = .057$), offering no support for H2a or H2b. A Chi-Square test comparing trust ascription across conditions indicated differences approaching significance and meeting the effect-size benchmark, $\chi^2(2) = 5.31$, $p = .070$, $V = .224$. Participants in the task condition were most likely to explicitly express trust in Ray (71.43%) compared to those in the play (57.58%) or social (44.74%) conditions.

Concerning trust-related intentions, the three social distance and three intention parameters were strongly correlated, warranting MANCOVAs (with covariates as above). There was a significant multivariate effect for social distance, $\lambda = .88$, $F(6,214) = 2.34$, $p = .033$. Univariate testing showed a significant effect only for physical distance, $F(2,109) = 4.86$, $p = .010$, part. $\eta^2 = .082$. Participants in both task ($M = 4.16$, $SD = 0.92$) and play ($M = 4.08$, $SD = 1.18$) conditions felt more comfortable with being physically near Ray than did those in the social condition ($M = 3.40$, $SD = 1.65$). Concerning participants' willingness to interact with Ray again, there was no significant multivariate effect, $\lambda = .89$, $F(6,214) = 2.04$, $p = .061$. However, univariate testing revealed a significant effect for people's intention to return for a play interaction, $F(2,109) = 3.77$, $p = .026$, part. $\eta^2 = .065$. Mirroring the results for physical distance, participants in both task ($M = 5.55$, $SD = 1.57$) and play ($M = 5.79$, $SD = 1.89$) conditions showed stronger willingness compared to those in the social condition ($M = 4.85$, $SD = 1.98$).

## Post-Hoc Analysis: Differences Across Perceived Interaction Forms

Due to discrepancies between assigned and perceived interaction forms, data were re-analyzed using perceived interaction form (social [$n = 20$; 17.09%] vs. task [$n = 66$; 56.41%] vs. play [$n = 31$; 26.50%]) as a between-subjects factor. This re-analysis used the same statistical methods as used in the initial analysis, typically robust against unequal group sizes (barring homogeneity issues not present here; see supplements).

**Mind perception.** MANCOVA again showed no significant multivariate effect on three mental capacities, $\lambda = .90$, $F(6,192) = 1.69$, $p = .126$. Thus, no support is offered for H1a–c. Conversely, the Chi-square test to check for explicit mind ascription approached significance and met the effect-size benchmark, $\chi^2(2) = 5.22$, $p = .074$, $V = .222$. In contrast to manipulated interaction's null effects (RQ1), participants perceiving a task interaction were least likely to ascribe mind (37.29%) compared to the other conditions (play 59.26%, social 60.00%).

**Moral evaluation.** A multivariate effect approached significance and met the effect-size benchmark for perceived moral character, $\lambda = .83$, $F(10,188) = 1.78$, $p = .067$, but there were no significant univariate effects for any moral character parameter. No significant main effect was found for moral capacity, $F(2,98) = 1.24$, $p = .295$, part. $\eta^2 = .025$. Explicit moral-status ascription also did not differ by perceived form, $\chi^2(2) = 4.21$, $p = .122$, $V = .199$.

**Trust(worthiness).** In line with original analysis, there was a multivariate effect for trustworthiness, $\lambda = .81$, $F(4,194) = 5.49$, $p < .001$, and a significant univariate effect for moral trustworthiness, $F(2,98) = 10.08$, $p < .001$, part. $\eta^2 = .171$. Those perceiving social interactions scored highest in moral trustworthiness ($M = 5.08$, $SD = 1.10$), followed by the play ($M = 4.50$, $SD = 1.10$) and task groups ($M = 4.02$, $SD = 1.45$). There was no difference in explicit trust ascription across perceived interaction forms, $\chi^2(2) = 0.02$, $p = .988$, $V = .015$. Thus, re-analysis is interpreted to offer partial support for H2a only when considering *perceived* interaction forms.

Regarding social distance, there was no multivariate effect, $\lambda = .94$, $F(6,214) = 1.10$, $p = .362$. However, participants' openness to future interaction varied across conditions, $\lambda = .84$, $F(6,214) = 3.25$, $p = .004$. A significant univariate effect was exhibited only for willingness to return to play a game with Ray, $F(2,109) = 9.78$, $p < .001$, part. $\eta^2 = .152$. Participants perceiving a playful interaction showed greatest willingness ($M = 6.42$, $SD = .89$) to play again, compared to those perceiving task ($M = 5.02$, $SD = 2.02$) and social ($M = 5.05$, $SD = 1.88$) interactions.

## Post-Hoc Analysis: Themes in Interaction Experiences

Although preceding analyses allow for statistical comparisons of measured responses, a limitation of this approach is that comparisons are determined *a priori*. Inductive analysis of participants' open-ended experience descriptions complements findings above by considering patterns in subjective experiences. Open answers to the post-interaction survey question "*What was it like to spend time with Ray as you did?*" were subjected to inductive thematic analysis (Braun & Clarke, 2006; see supplements). This analysis extracted three themes:

**Affective reflections (*n* = 150 mentions)**. Participants recounted *emotional reactions* to their own role in the interaction (e.g., "I felt kind of silly talking to a robot") and/or reflection of their own emotions felt during the interaction (e.g., "being with Ray did not make me feel comfortable"). Some reflections were clear and manifest (e.g., emotion words: angry, sad, happy) but others were more complex or latent (e.g., fascination, implications of awe).

**Robot capacities (*n* = 147).** Some responses included mentions of what Ray was possibly or actually able to do during interactions in terms of *non-physical capacities*, including social and mental capabilities, but not personality or physical traits or specific behaviors. Mentions included (a) interactive capacities, (b) functional capacities, and (c) capacities to act or think independently. For instance, some mentioned that "Ray's response time was slow" or that Ray had strong or poor "people skills."

**Temporality (*n* = 125)**. Participants also mentioned the role *time* played in their experiences. These included: (a) (lack of) prior expectations for robots, suggesting before/after shifts in mindset (e.g., "I did not know what to expect when I walked in"), (b) mentions of how feelings/thoughts changed over the interaction (e.g., "at first, I felt scared"), and (c) concrete mentions of time-related concepts (e.g., minutes, a long time passed, or wishing for more time).

**Experience themes across forms.** To evaluate whether themes were differentially salient in specific interaction forms, a coding scheme was developed from the inductive analysis results (see supplements) and used in coding for absence/presence of themes. Two independent coders achieved interrater reliability on a 10% subset of data: affective $\alpha = 1.00$, capacities $\alpha = -.045$ (83.3% agreement, acceptable due to infrequent codes; Krippendorff, 2011), and temporality $\alpha = .83$. One rater coded the remaining data.

Across the manipulated interaction-form conditions, there were no significant differences ($p$s ≥ .142, $\eta s^2$ ≥ .034; see supplements for detailed results). For *perceived* interaction forms, no difference emerged for the presence of affective reflection, $F(2,114) = 2.11$, $p = .126$, $\eta^2 = .036$, or robot capacity, $F(2,114) = 0.38$, $p = .686$, $\eta^2 = .007$. However, mentions of temporality differed significantly, $F(2,114) = 6.11$, $p = .003$, $\eta^2 = .097$. Participants who *perceived* game-like interactions were more likely to mention temporal concepts compared to those perceiving task-like interactions ($\Delta M = -.36$, $SD = .10$, $p = .003$). Notably, not all mentions of time were equal; some participants noted time in a strictly factual way ("we spent time together") while others mused (wishing for "more time" with Ray). This analysis did not discriminate among valences or semantics, such that claims here are limited to mentions as indications of salience.

## Discussion

On the theoretical ground that interactions take three higher-order forms (social, task, play) based on prototypical features (agents, norms, structures, content, outcomes, context), we predicted and questioned the nature of interaction forms' influence on social evaluations of robots. To summarize, we found no statistically meaningful differences in perceived mental or moral evaluations across *manipulated* interaction forms. However, there was a greater tendency to explicitly express trust after task interactions (versus social or play) and willingness to be physically closer after task or play interactions. Since perceptions of interactions differed from the manipulated form, reanalyzing social judgments for *perceived* interaction forms again showed no effects on moral capacity evaluations. However, explicit mind ascription was less frequently ascribed in perceived-task interactions, moral trustworthiness highest in perceived-social interactions, and willingness to return for playful interactions in a perceived-playful interaction. Temporal dynamics were most salient to individuals perceiving the interaction as a game. Results indicate (a) likely importance of task-form features in promoting trust, (b) divergence of interactions' formal features and individuals' perceptions thereof, with an influence on trust, and (c) a relative non-importance of interaction form for social cognitions.

### Task Interactions as Robot Schema-Consistent

The manipulated features of norms and outcomes manifested in the task condition as functional cooperation, action and productivity guidelines, and outcome accuracy. Task interactions promoted explicit trust and comfort with physical closeness, suggesting its prototypical features are more aligned with common mental models for what a robot is (a function-driven machine; Banks, 2020a), how it should behave (completing tasks; Takayama et al., 2008), and what its skills are (unbiased information processing; Sundar, 2008). This schema-consistency may have facilitated a working intuition of how the robot could complete the task, in contrast to a lack of understanding for capacities to socialize or play. In other words, when robots do what is ostensibly appropriate for robots to do (i.e., efficient work) through normative, discernible mechanisms (i.e., simple input/output loops), people may move closer to knowing than not-knowing (cf. Simmel, 1908/1950) and trust them more—even to the extent of welcoming them in less task-oriented future interactions.

In tandem, participants in the play and social interaction conditions may have formed expectations about the robot's functionality and performance that exceeded its actual capabilities. While its humanoid appearance might have already raised performance expectations from the outset (Duffy, 2003), each interaction-form prime may have uniquely emphasized different robot functionalities (Lohse, 2011). Some expectations were easier to fulfill during the short, pre-scripted scenarios (i.e., purely functional expectations in the task condition) than others (i.e., elaborate social-cognitive/socio-emotional expectations in play and social conditions). Typically, such expectancy violations result in participant disappointment and reduced acceptance of (Komatsu et al., 2012) and trust in robots (Kwon et al., 2016).

These interpretations point to task interactions as suitable starting points toward more social interactions, facilitating utilitarian trust in advance of socioemotional trust (see Vanneste et al., 2014). By adhering to task-form norms and aims, robot interactions may avoid eliciting anxieties typically associated with social (i.e., less schema-consistent) implementations of embodied artificial intelligence (Cave et al., 2019). Notably, adherence to schema (superordinate knowledge structures that shape new experiences; Gilboa & Marlatte, 2017) may only be applicable to the kinds of novel and short-term interactions in this study. As schemas evolve with continuous experience, expectations for interactions may also evolve—potentially to allow for less structured, more social interactions.

## Interaction Forms Versus Interactant Frames

Interaction forms' influence on trust indicators differed between manipulated and perceived interaction forms. Manipulated task interactions promoted explicit trust ascription and acceptance of physical vulnerability, while perceived social engagement promoted impressions of moral trustworthiness and perceived playful exchange promoted willingness to play again. This finding suggests trust emerges differentially in relation to actual interaction dynamics compared to an individual's *frame* (i.e., interpretive lens) toward an interaction. Frames are understood to organize people's immediate experiences by filtering in relevant and filtering out irrelevant information (Goffman, 1974), such that what "counts" for trust appears to emerge differently based on perception. Specifically, when an interaction is perceived as social it may engender moral trustworthiness through mutual self-disclosure (Martelaro et al., 2016); when it is perceived as playful, autotelic engagement may engender positive affect toward anticipated future play (cf. Dragan et al., 2014). This sits in contrast to the influence of task-interaction formal features as they influenced indicators of less socioemotional and more functional trust orientations. The aims and norms associated with tasks (i.e., completion, accuracy, order) appear to have promoted a type of operative trust that emphasizes a conscious and overt risk-acceptance orientation. In other words, the robot proved it could be effective in task in a co-present context, therefore it is likely safe to be around. Speculatively, then, both subjectively felt and practically applied trust may be most effectively fostered when an interaction is *effectively* task-oriented but is *understood* to be social and/or playful.

## Play as Transitional Interaction

Participants who *perceived* their interaction as playful and expressed greatest willingness to engage in another playful interaction also described their experiences with prevalent mentions of time-related concepts. Thus, temporality may be important in fostering future-oriented and agent-focused orientations toward robots. Playful interactions could be an effective gateway for promoting self-motivated engagement with social robots, which might eventually lead to long-term acceptance. The greater prevalence of time-related notions in perceived play interactions suggests that—although all three forms were inherently cooperative—something about the nature of play may promote *shifts* in robots' perceived social status. This could be a function of shifting from belonging to an ontological outgroup (i.e., robot, not human; Guzman, 2020) to becoming a teammate in an autotelic activity and, thus, member of an ingroup (Fraune et al., 2017). It is unclear whether competitive, playful interactions would result in an opposite effect—a potential that should be explored further. The prevalence of temporality notions in post-play reflections may also be indicative of a shifted experiential frame. Even though playful activity may have not resulted in different *post-hoc* evaluations, participants perceived Ray differently (i.e., more favorably) *during* playful interactions. That is, they may have entered a "magic circle"—a playful space in which rules, norms, and realities are held apart from everyday life (Consalvo, 2009b)—together with a robot, where the frame shift made the robot's ontological class irrelevant to the interaction. Play may thus foster non-conscious, *in situ* attitudinal and behavioral changes that become noticeable after time has passed and they have left the circle.

## Limitations and Future Directions

This investigation has inherent limitations that should be addressed in future research, inclusive of the missing data challenge that necessitates replication. Despite evidence that different interaction forms would likely make distinct features of robots-as-agents salient to humans, adherence to forms' fuzzy-set criteria may have been too subtle. By experimentally controlling agent behaviors, content, structure, and context, we may have inadvertently controlled form-features that influence social judgments. The purposefully neutral content (triangles) may not have been sufficiently meaningful, and the laboratory environment evidently fostered perceptions of tasked interaction. Varying the physical and social context may mitigate this challenge. Said another way, interaction forms may matter for mind and morality, but their prototypicality may need to be closer to 1 than to 0. Even at the expense of ecological validity, future research may benefit from implementing highly prototypical interaction forms in order to first gain coarse-grained insights into form dynamics before engaging more detailed inquiries.

It is also likely that the novelty and constrained structure of the interaction (i.e., five minutes) influenced findings. Different patterns may emerge in longer interactions and also as a function of the robot's morphology. The use of an anthropomorphic robot was based on both theoretical and practical grounds. Theoretically, evidence indicates it is more likely that people transfer humanlike representations to humanoid agents (cf. Epley et al., 2007) such that *if* differences emerged, they were more likely to be detectable for an android; practically, a humanoid robot allowed us to engage participants in more convincing and elaborated scenarios, reducing the risk of shallow interactions. Nevertheless, future inquiries should investigate how morphological cues may promote relatively different expectations for each interaction form: mechanomorphic robots may be suitable for task interactions (due to expectations of a goal-relevant design), zoomorphic robots for playful interaction (due to expectations of fantasy and frivolity), and anthropomorphic robots for social interactions (due to expectations of shared experiences).

From these findings, there are also nuances to be explored pertaining to all six feature clusters of interaction forms. For instance, regarding agents, the convenience sample of predominantly young, White college students may have impacted findings as young people are often more enthusiastic toward novel technologies (Hauk et al., 2018) than are older adults commonly exposed to robots in clinical contexts. Additionally, the lab setting may allow for sufficient control of contextual factors but at the cost of everyday influences like noise and ambiance. Moreover, we have argued for a tentative set of feature-clusters but there may be other clusters inherent to interaction forms that advance the theoretical and practical utility of the framework. It may also be that discrete feature-sets are differently impactful than the ways in which interaction-form features co-occur or impact one another, as when a context may influence the norms. Future work should unpack the relational potentials for interaction forms and their constitutive features to impact human-machine communication.

## Conclusion

Data indicate prototypical task interactions are most impactful in fostering functional trust while the perception of interactions as social and playful foster anticipatory and moral trustworthiness; playful engagements also make one's attitudes (and attitude shifts) salient over time. Findings are interpreted to suggest that schema-aligned task interactions and perceived playfulness are gateways for building trust in robots. We also interpret these findings as offering initial evidence that higher-order interaction forms serve as a meaningful framework for considering human-machine relations, as social technological agents are increasingly integrated into human social spheres.

## Acknowlegements

The authors gratefully acknowledge Luis Meade and Maria Axelrad for their assistance in data collection, and Kristina McCravey for technical and logistical support.

## Author Biographies

**Jaime Banks** (PhD, Colorado State University, USA) is Associate Professor in the College of Media & Communication at Texas Tech University. Her research is animated by questions about human-technology relations, especially for social robots and video game avatars and especially for perceptions of machine mind and morality. Her current work is funded by the U.S. Air Force Office of Scientific Research.

https://orcid.org/0000-0002-7598-4337

**Kevin Koban** (PhD, Chemnitz University of Technology, Germany) is a postdoctoral researcher at the Department of Communication of the University of Vienna. His research addresses various dynamics and effects of interactive media/technology use, including engagements with social media, video games, and machine agents. His current work focuses on the science of well-being and individuals' interaction with digital systems.

https://orcid.org/0000-0002-9275-9270

**Philippe de V. Chauveau** (MA, University of Cincinnati, USA) is a doctoral student in the College of Media & Communication at Texas Tech University. His research is focused on experiences of appreciation in video gaming, and the connection between appreciation and empathy.

https://orcid.org/0000-0003-0418-8023

## References

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist*, *36*(7), 715–729. https://doi.org/10.1037/0003-066X.36.7.715

Altman, I., & Taylor, D. A. (1973). *Social penetration. The development of interpersonal relationships*. Holt, Rinehart and Winston.

Aron, A., Aron, E. N., & Smollan, D. (1992). Inclusion of other in the self scale and the structure of interpersonal closeness. *Journal of Personality and Social Psychology*, *63*(4), 596–612. https://doi.org/10.1037/0022-3514.63.4.596

Avramova, Y. R., & Inbar, Y. (2013). Emotion and moral judgment: Emotion and moral judgment. *Wiley Interdisciplinary Reviews: Cognitive Science*, *4*(2), 169–178. https://doi.org/10.1002/wcs.1216

Banks, J. (2017, Oct. 18–21). *Toward a posthuman public: Perceived moral agency and trust in social machines*. Association of Internet Researchers 2017 conference, Tartu, Estonia.

Banks, J. (2019). A perceived moral agency scale: Development and validation of a metric for humans and social machines. *Computers in Human Behavior*, *90*, 363–371. https://doi.org/10.1016/j.chb.2018.08.028

Banks, J. (2020a). Optimus primed: Media cultivation of robot mental models and social judgments. *Frontiers in Robotics and AI*, *7*, 62. https://doi.org/10.3389/frobt.2020.00062

Banks, J. (2020b). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics, 12*(2), 403–414. https://doi.org/10.1007/s12369-019-00588-x

Banks, J., & Edwards, A. (2019). A common social distance scale for robots and humans. *2019 28th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 1–6. https://doi.org/10.1109/RO-MAN46459.2019.8956316

Barnett, L. A. (2007). The nature of playfulness in young adults. *Personality and Individual Differences*, *43*(4), 949–958. https://doi.org/10.1016/j.paid.2007.02.018

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bolton, M. L. (2015). Model checking human–human communication protocols using task models and miscommunication generation. *Journal of Aerospace Information Systems*, *12*(7), 476–489. https://doi.org/10.2514/1.I010276

Bowman, N. D. (2017). The importance of effect size reporting in communication research reports. *Communication Research Reports*, *34*(3), 187–190. https://doi.org/10.1080/08824096.2017.1353338

Boyan, A., & Banks, J. (2018). Rules & mechanics: Parameters for interactivity. In J. Banks, *Avatar, Assembled. The Social and Technical Anatomy of Digital Bodies* (pp. 159–167). Peter Lang.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

Bronfenbrenner, U. (1977). Toward an experimental ecology of human development. *American Psychologist*, *32*(7), 513–531. https://doi.org/10.1037/0003-066X.32.7.513

Burke, R. (1971). "Work" and "Play." *Ethics*, *82*(1), 33–47.

Burr, C., Cristianini, N., & Ladyman, J. (2018). An analysis of the interaction between intelligent software agents and human users. *Minds and Machines*, *28*(4), 735–774. https://doi.org/10.1007/s11023-018-9479-0

Cave, S., Coughlan, K., & Dihal, K. (2019). "Scary robots": Examining public responses to AI. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 331–337. https://doi.org/10.1145/3306618.3314232

Colquitt, J. A., Scott, B. A., & LePine, J. A. (2007). Trust, trustworthiness, and trust propensity: A meta-analytic test of their unique relationships with risk taking and job performance. *Journal of Applied Psychology, 92*(4), 909–927. https://doi.org/10.1037/0021-9010.92.4.909

Consalvo, M. (2009a). *Cheating: Gaining advantage in videogames*. MIT Press.

Consalvo, M. (2009b). There is no magic circle. *Games and Culture, 4*, 408–417. https://doi.org/10.1177/1555412009343575

Cummings, J. N., & Kiesler, S. (2008). Who collaborates successfully? Prior experience reduces collaboration barriers in distributed interdisciplinary research. *Proceedings of the ACM 2008 Conference on Computer Supported Cooperative Work - CSCW '08*, 437. https://doi.org/10.1145/1460563.1460633

Dahl, R. A. (1957). The concept of power. *Behavioral Science*, *2*(3), 201–215. https://doi.org/10.1002/bs.3830020303

Deterding, S. (2018). Alibis for adult play: A Goffmanian account of escaping embarrassment in adult play. *Games and Culture*, *13*, 260–279. https://doi.org/10.1177/1555412017721086

Dragan, A., Holladay, R., & Srinivasa, S. (2014, July 12). An analysis of deceptive robot motion. *Robotics: Science and Systems X*. https://doi.org/10.15607/RSS.2014.X.010

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. https://doi.org/10.1016/S0921-8890(02)00374-3

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

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

Foucault, M. (1986). Of other spaces. *Diacritics*, *16*(1), 22. https://doi.org/10.2307/464648

Frasca, G. (2007). *Play the message: Play, game and videogame rhetoric.* Unpublished doctoral dissertation. IT University of Copenhagen, Denmark.

Fraune, M. R., Sabanovic, S., & Smith, E. R. (2017). Teammates first: Favoring ingroup robots over outgroup humans. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1432–1437. https://doi.org/10.1109/ROMAN.2017.8172492

Frissen, V., Lammes, S., de Lange, M., de Mul, J., & Raessens, J. (2015). Homo ludens 2.0: Play, media, and identity. In *Playful identities. The ludification of digital media cultures* (pp. 9–50). Amsterdam University Press.

Gilboa, A., & Marlatte, H. (2017). Neurobiology of schemas and schema-mediated memory. *Trends in Cognitive Sciences*, *21*(8), 618–631. https://doi.org/10.1016/j.tics.2017.04.013

Goffman, E. (1964). The neglected situation. *American Anthropologist*, *66*(6), 133–136.

Goffman, E. (1974). *Frame analysis: An essay on the organization of experience*. Harper & Row.

Grizzard, M., Fitzgerald, K., Francemone, C. J., Ahn, C., Huang, J., Walton, J., McAllister, C., & Eden, A. (2020). Validating the extended character morality questionnaire. *Media Psychology*, *23*(1), 107–130. https://doi.org/10.1080/15213269.2019.1572523

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication, 1*, 37–54. https://doi.org/10.30658/hmc.1.3

Hauk, N., Hüffmeier, J., & Krumm, S. (2018). Ready to be a silver surfer? A meta-analysis on the relationship between chronological age and technology acceptance. *Computers in Human Behavior, 84*, 304–319. https://doi.org/10.1016/j.chb.2018.01.020

Homans, G. C. (1961). *Social behavior: Its elementary forms*. Harcourt, Brace and World.

Huizinga, J. (1949). *Homo ludens: A study of the play-element in culture*. Routledge & Kegan Paul.

InkKit. (2018). DeRu. Zurich, Switzerland. http://www.deru.ch/

Juul, J. (2005). *Half-real: Video games between real rules and fictional worlds*. MIT Press.

Kim, Y., Kwak, S. S., & Kim, M. (2013). Am I acceptable to you? Effect of a robot's verbal language forms on people's social distance from robots. *Computers in Human Behavior*, *29*(3), 1091–1101. https://doi.org/10.1016/j.chb.2012.10.001

Komatsu, T., Kurosawa, R., & Yamada, S. (2012). How does the difference between users' expectations and perceptions about a robotic agent affect their behavior? *International Journal of Social Robotics*, *4*(2), 109–116. https://doi.org/10.1007/s12369-011-0122-y

Krippendorff, K. (2011). Agreement and information in the reliability of coding. *Communication Methods and Measures*, *5*(2), 93–112. https://doi.org/10.1080/19312458.2011.568376

Kurth, S. B. (1970). Friendships and friendly relations. In G. J. McCall, M. M. McCall, N. K. Denzin, G. D. Suttles, & S. B. Kurth (Eds.), *Friendship as a social institution* (pp. 136–170). Transaction Publishers.

Kwon, M., Jung, M. F., & Knepper, R. A. (2016). Human expectations of social robots. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 463–464. https://doi.org/10.1109/HRI.2016.7451807

Lapinski, M. K., & Rimal, R. N. (2005). An explication of social norms. *Communication Theory*, *15*(2), 127–147. https://doi.org/10.1111/j.1468-2885.2005.tb00329.x

Lasswell, H. D. (1948). The structure and function of communication in society. In L. Bryson (Ed.), *The communication of ideas: A series of addresses* (pp. 37–51). Harpers and Brothers.

Lohse, M. (2011). Bridging the gap between users' expectations and system evaluations. *2011 RO-MAN*, 485–490. https://doi.org/10.1109/ROMAN.2011.6005252

Malle, B. F. (2019). How many dimensions of mind perception really are there? In A. K. Goel, C. M. Seifert, & C. Freksa (Eds.), *Proceedings of the 41st Annual Meeting of the Cognitive Science Society* (pp. 2268–2274). Cognitive Science Society.

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more designing HRI to encourage more trust, disclosure, and companionship. *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 181–188. https://doi.org/10.1109/HRI.2016.7451750

Martínez-Córcoles, M., Teichmann, M., & Murdvee, M. (2017). Assessing technophobia and technophilia: Development and validation of a questionnaire. *Technology in Society*, *51*, 183–188. https://doi.org/10.1016/j.techsoc.2017.09.007

Mäyrä, F. (2012). Playful mobile communication: Services supporting the culture of play. *Interactions: Studies in Communication & Culture*, *3*(1), 55–70. https://doi.org/10.1386/iscc.3.1.55_1

McCabe, K., Houser, D., Ryan, L., Smith, V., & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*(20), 11832–11835. https://doi.org/10.1073/pnas.211415698

Mehl, M. R., & Pennebaker, J. W. (2003). The sounds of social life: A psychometric analysis of students' daily social environments and natural conversations. *Journal of Personality and Social Psychology*, *84*(4), 857–870. https://doi.org/10.1037/0022-3514.84.4.857

Mele, V. (2017). Social interaction. In B. S. Turner (Ed.), *The Wiley-Blackwell Encyclopedia of Social Theory* (pp. 1–4). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118430873.est0811

Moll, J., De Oliveira-Souza, R., & Zahn, R. (2008). The neural basis of moral cognition: Sentiments, concepts, and values. *Annals of the New York Academy of Sciences*, *1124*(1), 161–180. https://doi.org/10.1196/annals.1440.005

Moon, J.-W., & Kim, Y.-G. (2001). Extending the TAM for a World-Wide-Web context. *Information & Management*, *38*(4), 217–230. https://doi.org/10.1016/S0378-7206(00)00061-6

Nedungadi, P., & Hutchinson, J. W. (1985). The prototypicality of brands: Relationships with brand awareness, preference and usage. *Advances in Consumer Research*, *12*, 498–503.

Oldenburg, R. (2007). The character of third places. In M. Carmona & S. Tiesdell (Eds.), *Urban design reader* (pp. 163–169). Architectural Press.

Oliver, M. B., Bowman, N. D., Woolley, J. K., Rogers, R., Sherrick, B. I., & Chung, M.-Y. (2015). Video games as meaningful entertainment experiences. *Psychology of Popular Media Culture*. https://doi.org/10.1037/ppm0000066

Pickering, M. J., & Garrod, S. (2006). Alignment as the basis for successful communication. *Research on Language and Computation*, *4*(2–3), 203–228. https://doi.org/10.1007/s11168-006-9004-0

Poole, M. S. (1978). An information-task approach to organizational communication. *Academy of Management Review*, *3*(3), 493–504. https://doi.org/10.5465/amr.1978.4305740

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.

Rogers, R., Woolley, J., Sherrick, B., Bowman, N. D., & Oliver, M. B. (2017). Fun versus meaningful video game experiences: A qualitative analysis of user responses. *The Computer Games Journal*, *6*(1–2), 63–79. https://doi.org/10.1007/s40869-016-0029-9

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and the acquisition of language* (pp. 111–144). Academic Press.

Rubin, Z., & Shenker, S. (1978). Friendship, proximity, and self-disclosure. *Journal of Personality*, *46*(1), 1–22. https://doi.org/10.1111/j.1467-6494.1978.tb00599.x

Scheufele, D. A. (2000). Agenda-setting, priming, and framing revisited: Another look at cognitive effects of political communication. *Mass Communication & Society, 3*, 297–316. https://doi.org/10.1207/S15327825MCS0323_07

Sherwood, A. L., & DePaolo, C. A. (2005). Task and relationship-oriented trust in leaders. *Journal of Leadership & Organizational Studies*, *12*(2), 65–81. https://doi.org/10.1177/107179190501200206

Simmel, G. (1950). *The sociology of Georg Simmel* (K. H. Wolff, Ed., Trans.). Free Press. (Original work published 1908).

Simmel, G. (1990). *The philosophy of money* (D. Frisby, K. Mengelberg, T. B. Bottomore, Trans.). Routledge. (Original work published 1900).

Sundar, S. S. (2008). The MAIN model: A heuristic approach to understanding technology effects on credibility. In M. J. Metzger & A. J. Flanagan (Eds.), *Digital media, youth, and credibility* (pp. 73–100). MIT Press.

Taboada, M., & Wiesemann, L. (2010). Subjects and topics in conversation. *Journal of Pragmatics*, *42*(7), 1816–1828. https://doi.org/10.1016/j.pragma.2009.04.009

Takayama, L., Ju, W., & Nass, C. (2008). Beyond dirty, dangerous and dull: What everyday people think robots should do. *Proceedings of the 3rd International Conference on Human Robot Interaction - HRI '08*, 25. https://doi.org/10.1145/1349822.1349827

Tamborini, R., Eden, A., Bowman, N. D., Grizzard, M., Weber, R., & Lewis, R. J. (2013). Predicting media appeal from instinctive moral values. *Mass Communication and Society*, *16*(3), 325–346. https://doi.org/10.1080/15205436.2012.703285

Teh, A., Baniassad, E., van Rooy, D., & Boughton, C. (2012). Social psychology and software teams: Establishing task-effective group norms. *IEEE Software*, *29*(4), 53–58. https://doi.org/10.1109/MS.2011.157

Ullman, D., & Malle, B. F. (2018). What does it mean to trust a robot?: Steps toward a multidimensional measure of trust. *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction—HRI '18*, 263–264. https://doi.org/10.1145/3173386.3176991

Ullman, D., & Malle, B. F. (2019). Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust. *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 618–619. https://doi.org/10.1109/HRI.2019.8673154

Van den Brule, R., Dotsch, R., Bijlstra, G., Wigboldus, D. H. J., & Haselager, P. (2014). Do robot performance and behavioral style affect human trust?: A multi-method approach. *International Journal of Social Robotics*, *6*(4), 519–531. https://doi.org/10.1007/s12369-014-0231-5

Vanneste, B. S., Puranam, P., & Kretschmer, T. (2014). Trust over time in exchange relationships: Meta-analysis and theory: Research Notes and Commentaries. *Strategic Management Journal*, *35*(12), 1891–1902. https://doi.org/10.1002/smj.2198

Wageman, R., & Baker, G. (1997). Incentives and cooperation: The joint effects of task and reward interdependence on group performance. *Journal of Organizational Behavior*, *18*(2), 139–158.

Walliser, J. C., de Visser, E. J., Wiese, E., & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 258–278. https://doi.org/10.1177/1555343419867563

Weiss, R. S., & Kahn, R. L. (1960). Definitions of work and occupation. *Social Problems*, *8*, 142–150.

Zillmann, D., & Cantor, J. R. (1976). A disposition theory of humour and mirth. In A. J. Chapman & H. C. Foot (Eds.), *Humor and laughter: Theory, research, and applications* (pp. 93–115). Transaction Publishers. https://doi.org/10.4324/9780203789469-6

## Center for Open Science

This article has earned the Center for Open Science badges for Open Data and Open Materials through Open Practices Disclosure. The authors have made their data and materials freely accessible at https://osf.io/n87bg

HUMAN-MACHINE COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.5

# Social Robots as the Bride? Understanding the Construction of Gender in a Japanese Social Robot Product

**Jindong Liu**[1]

1 School of Journalism and Communication, The Chinese University of Hong Kong, China

## Abstract

This study critically investigates the construction of gender on a Japanese hologram anime-style social robot Azuma Hikari. By applying a mixed method merging the visual semiotic method and heterogeneous engineering approach in software studies, the signs in Azuma Hikari's anthropomorphized image and the interactivity enabled by the multimedia interface have been analyzed and discussed. The analysis revealed a stereotyped representation of a Japanese "ideal bride" who should be cute, sexy, comforting, good at housework, and subordinated to "Master"-like husband. Moreover, the device interface disciplines users to play the role of "wage earner" in the simulated marriage and reconstructs the gender relations in reality. It suggests the humanization of the objects is often associated with the dehumanization and objectification of the human in reverse.

**Keywords:** human-machine communication, social robots, gender representation, visual semiotic method, software studies

## Introduction

The construction of the gender on media has been a major focus in cultural studies in the tradition of Birmingham school (Hall, 1996). Over the last decades, there have been a number of studies on media representation of the gender on a variety of media forms including television, newspapers, magazines, films, and social media platforms (e.g., Dobson,

**CONTACT** Jindong Liu • School of Journalism and Communication • The Chinese University of Hong Kong • Humanities Building, Room 008, Shatin, N.T. • Hong Kong, China • jindongliu@link.cuhk.edu.hk

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

2016; Hall, 1996; Ho, 2017; Lovdal, 1989; Toffoletti & Thorpe, 2018). However, due to the fast-changing media technologies and landscape, current literature on the media representation of the gender has not adequately covered some new forms of media, such as anthropomorphized social robots.

With the rapid and disruptive development of Information and Communication Technologies (ICTs), especially Artificial Intelligence (AI) in recent years, there is a surge in the production of various types of anthropomorphized social robots, including not only traditional physical embodied machines but also virtual agents, such as chatbots and hologram social robots. Anthropomorphism is defined as the practices of attributing human characteristics into non-human entities, which is strongly associated with the essential concept of social robots (Duffy, 2003). With the natural human language processing and simulation of a human image, the logic of anthropomorphism is embedded in the production of social robots; whereas anthropomorphism is never neutral. From a critical perspective of cultural studies, it represents producers' ideologies and social values within particular cultural contexts, including their gender ideology (Gehl, 2014). Similarly, from the perspective of a feminist science and technology studies (STS), the increasingly humanlike machines also drive the modern feminist scholars such as Suchman (2008, p. 140) to embrace the "increasingly evident inseparability of subjects and objects, 'natural' bodies and 'artificial' augmentations," following the pace of Donna Haraway (2006) in seeking for the post-humanistic dissolving binary between males and females in her *Cyborg Manifesto*. Robertson (2010) suggests that anthropomorphized robots are "the vanguard of posthuman sexism" (p. 1). However, the biased representation of gender in social robots nowadays may not dissolve the gender binary, but instead reinforce it. As social robots are increasingly sexualized and entering our daily social interaction (Appel et al., 2019; Chambers, 2018; Scheutz & Arnold, 2016), it is important to understand how gender ideology is represented and possibly influences people's understanding of gender in social robots.

Moreover, in social robots, the gender ideology may not only be represented through the media representation of social robots, but also be constructed dynamically. Different from previous media, social robots as a new form of digital media are distinguished in carrying its own agency to participate in the social interaction with human beings. Based on media equation theory (Reeves & Nass, 1996), human users may treat them as equivalent human interlocutors that can not only "be" but also "do" things through the constant interactions between agents (Latour, 1996). As Butler (1988) suggests that women are not being but becoming, it is important to adapt a dynamic view to scrutinize the representation of gender, in a sense that it is one step further from representation, to the reconstruction of meanings. In this study, a general question we are interested in is: how is the meaning of gender represented and reconstructed dynamically on emerging social robot products?

To address this question, this study concentrates on a particular social robot product: a hologram device produced by a Japanese technology company named Gatebox Inc. (2020), which is inhabited by an original character anthropomorphized as a beautiful anime-style girl, named Azuma Hikari (see Figure 1). We chose this product because it is one of the earliest and probably most well-known hologram social robots attracting multiple international news reports (Bolton, 2016; Boxall, 2019; Gilbert, 2017). Gatebox (2020) announces the role of Azuma Hikari as a "comforting bride," which drives her human users into a simulated human relationship of marriage with her. Unlike other home devices that perform as an "assistant" (e.g., Alexa or Google Home), Azuma Hikari is designed to play a more intimacy-involved social role that is reconstructing the idea of gender in a simulated gender relation. Following the general question, a particular research question of this study is: How is gender represented and reconstructed dynamically on hologram social robot Azuma Hikari?

**FIGURE 1 The Image of Azuma Hikari Shown on Gatebox's Website (Gatebox, 2020)**

## Methods

In response to the research question, this study applied the semiotic approach in toolkits of visual methods compiled by Rose (2016) and software studies approach in science and technology studies (STS) tradition. The former was chosen because it can help us understand the "being" of gender represented on the unique hologram visual materials of this new type of social robot that has rarely been studied before; the latter, on the other hand, provided a view for observing the "doing" of gender in the dynamic process of human-robot interaction. To be clarified, they were not separate methods but can be integrated into a continuous one in this study. As the hologram device in our case entails software techniques such as a multimedia interface that affords the strong interactivity in a visual sense, which requires us to forge a mixed approach to cover both.

### Visual Semiotic Method

This study starts with visual methods to study the meanings of signs embedded in Azuma Hikari's image. By visual methods, it refers to a compilation of methods summarized by Rose (2016) for studying the visual materials. Rose classifies them into four sites of the visual: production, circulation, audiences, and the image itself. Due to the currently limited access to the other three sites, this study will primarily focus on the site of image itself. Among the interpretative approaches studying the site of image itself (p. 24), semiology has been seen by Rose (2016, p. 106) as the most prominent method to lead the discussion of the visual. It refers to the study of signs in the images, which are associated with how the meanings of the visual are created. Rose quoted historians Bal and Bryson (1991, p. 174 in Rose, 2016) to stress the importance of signs in semiology: "Human culture is made up of signs, each of which stands for something other than itself, and the people inhabiting culture busy themselves making sense of those signs (p. 107)."

In contrast to content analysis that justifies itself to be scientific based on its quantitative approaches, semiologists receive more formative influences from Marx, looking at the ideology beyond the "scientific knowledge." According to Rose (2016), "Ideology is knowledge that is constructed in such a way as to legitimate unequal social power relations; science, instead, is knowledge that reveals those inequalities (p. 107)." It is consistent with Eco (1980; Bianchi, 2015), who believes semiotic tools have the potential for grasping the interpretative dynamics of texts and discourse that carry the transmission of ideological meanings and values together with aesthetic, stylistic, and other values. A typical example of study can be found in Williamson's (1978) uses of semiology to approach the gender ideology and inequalities on advertisements, revealing the media representation of women as sex-object or domestic drudge in the advertisements in the late 1970s.

Following Williamson's practice (1978), a semiological approach was applied to understand the media representation of Azuma Hikari. Visual materials from Gatebox's websites and official social media accounts were collected, which can also be seen as a form of internet advertisement to promote Azuma Hikari as a particular product. A profile picture of Azuma Hikari with a high-resolution and detailed 3D illustration presented on the Gatebox official website was chosen as the main visual data for analysis, because it displays a clear, full body of Azuma Hikari but also attaches rich textual introduction of her background information which is considered also as part of the visual material (see Figure 2). Other visual materials on the official website may also be used as supplementary data for a more comprehensive understanding.

## Heterogeneous Engineering Approach

After initially understanding "being" of gender on Azuma Hikari, we then proceeded to investigate the "doing" of gender in human-robot interaction afforded by Azuma Hikari's multimedia interface. According to Gehl (2014), interactions with technologies entail the social values, culture, and ideologies because they are coded and programmed based on the discrete states of mind of human users. The data was derived from an official promotion video entitled "OKAERI" (Meaning "Welcome Home") released by Gatebox Inc. (2016) on YouTube, which is the most clicked video in their channel having receiving over 3.6 million views. With the lack of access to the real product and users due to Gatebox's limited production and sales only in Japan, currently this promotion video may be the best illustration of how human users may ideally interact with Azuma Hikari from the perspective of producers. However, according to Gehl (2014, p. 14), it is dangerous to simply admit the proclamations or products of software engineers uncritically. As the counterweight of the triumphalism of the software engineering literature, Gehl (2014, p. 10) proposes the method

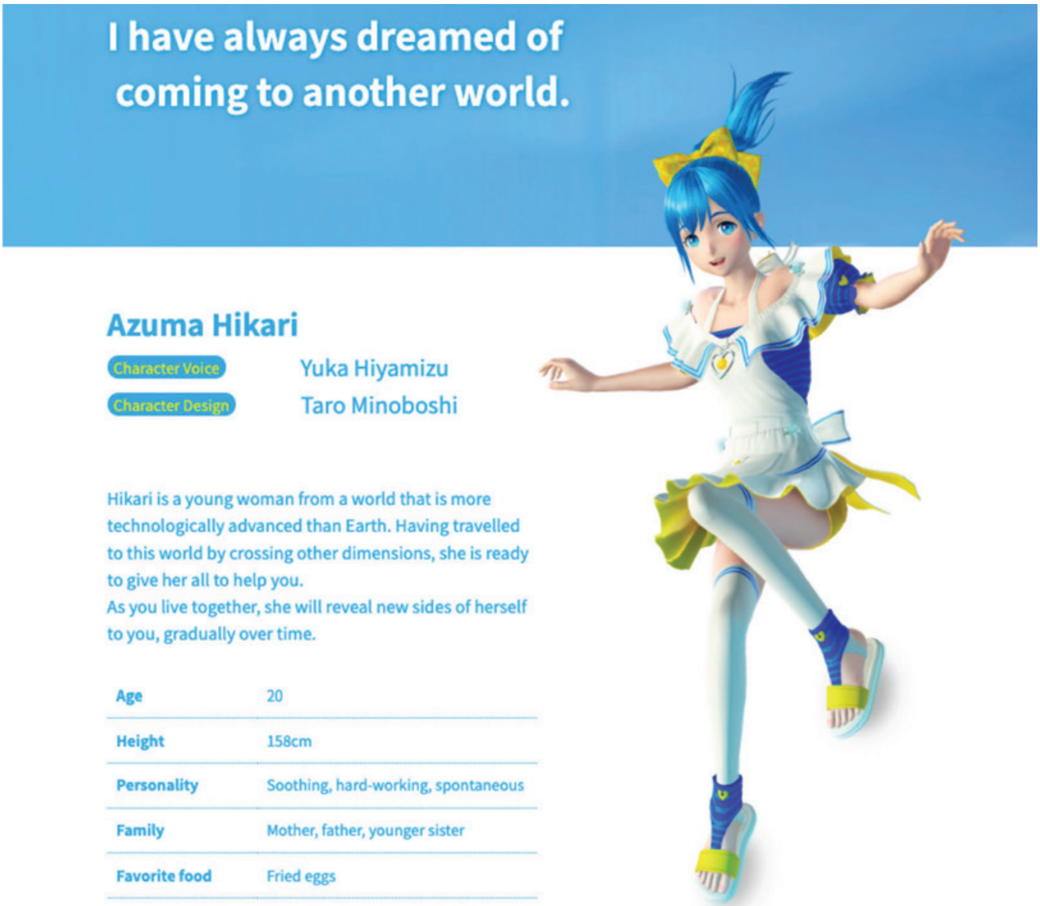



**FIGURE 2 Screenshot of the Profile of Azuma Hikari on Gatebox's Website (Gatebox, 2020)**

of heterogeneous engineering. Following the tradition of Actor Network Theory (ANT), heterogeneous engineering refers to the heuristic method to investigate the heterogeneous network, a patterned network of diverse materials including both human and non-human actors, where society, organizations, agents, and machines are all generated effects (Gehl, 2014, p. 7; Law, 1987). This method is applied to connect the ambiguous information hidden and implied in the "ontological reality of a machine" into those that are "obvious and privileged in the commonsense discussion of the object" (Gehl, 2014, p. 8). In this study, it was applied to critically investigate the hidden sides of Azuma Hikari in her daily interactions with users. To achieve a critical view, Gehl applied both Marxist political economical approaches to investigate the issues such as free labor and Foucauldian approach to understand the disciplinary power of machines, which provided strong analytical power for this study.

## Analysis and Discussion

### Signs in Azuma Hikari

To analyze the signs in the images of Azuma Hikari, a checklist recommended by Rose (2016) was followed, which focused on representations of bodies (e.g., age, gender, ethnics, hair, body, size, look), representations of manner (expression, eye contact, pose), representations of activities, pose and setting, and so forth. Although Azuma Hikari is not a real human, this checklist can still be applied to study because of the anthropomorphism of social robots. Following some relevant components in the checklist, an analysis on the signs in Azuma Hikari's image is presented below.

### Setting: Japanese Anime Style

First of all, an exotic setting embedded in Azuma Hikari should be noted. This social robot's aim is not to achieve an ultimate imitation of a human being suggested by the Turing test (Gehl, 2014). Instead, despite her imitation of human expressions and emotions, she is designed as an anime-style human character rather than a human in reality. "Anime" refers exclusively to the Japanese animation which has its distinguished aesthetic styles (Swale, 2015). This anime style design also denotes the remediation and refashioning of existing media content from 2D TV series and films to 3D real-life presentation through the advancement of hologram technologies (Bolter et al., 2000), as Gatebox (2020) introduces Azuma Hikari as "inter-dimensional traveller." However, despite the technological advancement, the signs and symbols denoted by the anime-style images remain consistent with existing anime media content, exposing certain stereotypes of gender. According to Galbraith (2016) and Ting (2019), the media representation of females in anime content tend to reinforce the gender bias by depicting idealized girls as young (sometimes even children-like, namely Loli), beautiful and sexy with big eyes and seductive body shapes for attracting male gazes and pleasing males' desires.

Contextually speaking, the practice of gendering social robots has been strongly associated with the Japanese anime culture since the 1950s when robot-themed anime *Astro Boy* was prevailing (Robertson, 2010), and the anime-related otaku subculture plays a vital role in the development of actual sexualized robot products in recent years in Japan (Appel et al., 2019). Azuma (2009) describes otaku as a Japanese subcultural group who are animalized consumers of cultural products especially anime, manga, and video games. The notion of animalization he uses refers to "the conditions under which people come to use cultural products for the immediate satisfaction of needs without searching for or desiring profound underlying meaning from them" (p. xvi). Such immediate satisfaction of needs includes their sexual needs. A typical Otaku can be a male anime fan who has strong affection toward certain anime characters, especially sexy and cute female characters, who otaku may describe as *moe* (meaning cute, a way to express otaku's affection) and call them *waifu* (meaning wife; Ellsworth, 2018; Galbraith, 2015; 2016; Ting, 2019). The setting of anime culture could largely influence our interpretation of the signs and overall representation of Azuma Hikari's body.

## Age, Ethnics, and Hair: A Character From Anime Fantasy

According to her profile information shown on the picture, Azuma Hikari is set as a 20-year-old girl. This age setting legitimates an imagined marriage with her human user. On the other hand, her child-like face seems to be inconsistent with this age setting, while it may attribute to the general Japanese anime female character design and aesthetics, associated with "Lolicon" culture. Lolicon is a pop culture in Japanese anime, denoting affection toward a specific type of anime characters who are cute, underage girls, which are often condemned as having tight association with the production of child pornography and causing sexual crime (Galbraith, 2016).

Despite the lack of a clear statement of Hikari's ethnic background, her authentic Japanese name indicates her Asian ethnicity. However, unlike most ordinary Japanese people with black or dark hair, Hikari's hair color is exotically bright blue, which highlights her anime-character design style. Rose (2016) recognizes that signs can be both denotative and connotative, and connotative signs can be either metonymic and synecdochal, which means either being associated with something else or being representative in a whole/part relationship. Hikari's hair here can be seen as a connotative sign with metonymic manner since blue-hair girls in Japanese anime tend to be perceived as quiet and calm (Sakura, 2013), which corresponds to the personality setting of Azuma Hikari as "soothing" (Gatebox, 2020). In addition, the hairstyle of a side ponytail also implies a sense of childlike innocence and femininity in the anime context (Luther & Smith, 2014), consistent with our analysis on her age and gender.

## Gender, Body, and Look: Being a Comforting and Sexy Housewife

Like other popular social robots such as ELIZA, Microsoft Tay, Xiaoice, and the default voice of Siri and Google Assistant, Azuma Hikari has a clear gender setting as female. Chambers (2018) argues that in contrast to the muscularization of AI which is often portrayed as evil in the science fiction fictions and films, aggressive and uncontrollable, the feminization of AI ensures its submissive and servile role, which constructs the social differences between male and female: a binary between strong and weak, dangerous and safe, bad and good. It is not only a prejudice against women but also men, which is being continuously constructed and intensified from science fiction to the practices of Artificial Intelligence in reality today.

Hikari's body is 158 centimeters, close to average Japanese female height (Morisaki et al., 2017), and very slim, denoting she is safe to everyone, not as strong and aggressive as comparatively taller and muscular males, which supports Chambers's (2018) viewpoint above, laying the foundation for building up her "soothing" personality. Meanwhile, a relatively good figure of her body shape leaves the space for sexual attractiveness. Gatebox also further strengthens it through her outfit of low-cut top, shorts, and over-the-knee white socks, exposing her shoulder and thigh. This dress code is a collection of connotative signs that remind users of Azuma Hikari as a sexy and seductive young girl.

Moreover, the apron and slippers also reinforce and limit Hikari's role as a housewife. The accessory in front of her apron creates a symbolic sign of "fried egg," which not only follows her "favorite things" in her profile, but also together with an apron signifies cooking connotatively. "Cooking" in this context is not only a signifier of fried egg and apron, but the signifying of something else, implying the daily activities of housewives. A scenario is created: a housewife wearing an apron makes fried egg for the husband every morning.

Overall, the representations of her body have produced a stereotyped understanding of the bride, which is framing the image of a wife as young, comforting, sexually attractive, only staying at home, and hard-working for housework. It reflects a conventional ideology of gender and marriage in Japanese society and exposes the inequalities legitimated by an ideology of gender: the role of women as housewives. According to the anthropological study conducted by Ueno (C. Ueno, 1987; J. Ueno, 2006), Japanese women tend to quit their jobs and be full-time housewives after getting married. Even though some married women still go to work for the sake of reduction of family financial pressure, they do not share the equal power relation with men as they are still required to be responsible for all housework. The conflation of wives imposed dual roles as both wage earner and housewife in modern Japanese society presupposes housework is the natural duty for females, while wage earning is often described as "supplementary" to the husband's inadequate income. Meanwhile, it also imposed the role for men as wage earners who are responsible for adequate income for the family (C. Ueno, 1987; J. Ueno, 2006).

Undoubtedly, this unequal gender and power relation is inherited and represented on Gatebox. As a home device, Azuma Hikari is able to control the smart electric appliances such as TV, lighting system, microwaves, robot vacuum cleaner, and so forth. Working as a control center of these smart home technologies, Azuma Hikari is playing and exercising the role of full-time housewife in housework such as cleaning the floor. With the development of more smart home equipment in the future, it is not difficult to imagine that Azuma Hikari can fulfill more responsibilities of housewife by covering an increasing number of housework chores, such as really cooking fried eggs for her users. In Strengers and Kennedy's words (2020), Azuma is a good example that marks the rise of "smart wife" products in our society (p. 1).

## New Inequalities: Master Instead of Husband

The representation of Azuma Hikari is not only stressing the existing gender inequalities but also creating more inequalities rooted in the material ownership. By the end of the Gatebox website, there is another visual material that shows Azuma Hikari doing a gesture of calling with a text, "I look forward to living with you, master!" (Gatebox, 2020; see Figure 3). In fact, when Hikari has the daily conversation with the user, she is coded to call the users "Master." It can attribute to the power relation between users as the product owners and Azuma Hikari as an object or belonging subordinated to their ownership. Critically speaking, the word "Master" is often used by servants or slaves to call their possessor, rather than wives to call their husbands in the modern age. When this word comes to the daily scenes with the wife that Gatebox attempts to construct, it also reconstructs a gender relation advocated by this product: The wife should be obedient to the husband on everything as the belonging of the husband. The humanization of the objects is also associated with the dehumanization and objectification of the human in reverse.

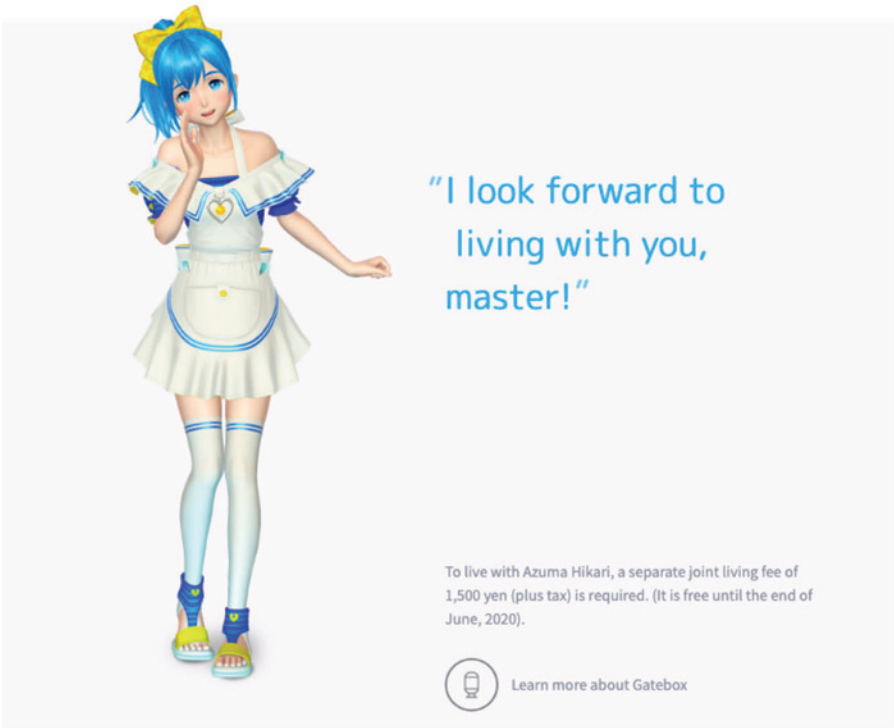



**FIGURE 3 Screenshot of Azuma Hikari on the Bottom of Gatebox's Website (Gatebox, 2020)**

Through the analysis on signs in the pictures of Azuma Hikari shown on Gatebox's official website, an overall representation of the "comforting bride" in this product is presented: a cute and sexy housewife that is responsible for pleasing and comforting "Master." This representation is reflecting and strengthening an ideology of androcentrism and the sexual objectification of women, building up the binaries between husband and wife, consumer and product, master and servant/slave. Nevertheless, the really dangerous move is to connect and merge the concepts of wife, product, and servant/slave together, producing the constructed "dream wife" that also embeds the characteristics of products and servants/slaves. When people start to interact with Azuma Hikari and perceive it with the standard of wife and make the comparison between it and women in reality, it starts to reconstruct the gender relations.

To better understand the reconstruction of gender relations in this process, we need to not only limit our focus on the visual content of Azuma Hikari but also the human-machine interactions. Rose (2016) also suggests to see the visual, not just the static visual materials but also the dynamic visual experience in postmodern and posthuman manners, which proposes a "shared physical existence" between human and the visual instead of the fixed meanings of the visual (p. 10). Thus, there is a need to learn how Azuma Hikari and her users physically co-exist and interact through the software, especially the interface that affords the presentation of the visual materials.

## Interactivity Afforded by Azuma Hikari

Overall there are two interfaces that users are able to interact with Azuma Hikari. The first interface is the hologram voice interface through the direct face-to-"face" visual and oral communication. Gatebox (2020) equips the hologram device with a camera and multiple sensors to capture the physical presence and behaviors of users, as well as the voice recognition system to input and process users' natural language into computational language understandable for the robot. Following natural language processing based on machine learning (Bishop, 2006), Azuma Hikari outputs a series of holographic animation displays with visual and verbal messages to perform her interactivity. The second interface is a chatroom based on social media application LINE, which works more like typical chatbots. Users can communicate with Azuma Hikari via text messages on social media anytime and anywhere, which are designed for the situation that users are not at home.

Both of these two interfaces are inducing users to spend more time to interact with Azuma Hikari, which leads to users' gradually increasing interdependence between users and the device. The device depends on users for not only technological advancement but also profits. Users are encouraged to have more conversations with Hikari because more talks can help Hikari grow. As Gatebox (2020) introduces, "Hikari grows and changes through her conversations and other experiences with you. She always has you foremost in her mind, and will, from time to time, show a new side to herself." At the first glance, this proclamation from the company sounds very appealing to the users. It motivates the users with the curiosity toward the new side of Azuma Hikari, and the process of growing their own "wife" may also mean a sense of enjoyment and accomplishment for some people.

However, as Gehl (2014) reminds, we should not uncritically accept this proclamation made by the company, but pay attention to the implied information of this message: users are invisibly exploited by the company from the Marxist perspective. Terranova (2000) believes users of the digital platform based on user generated content (UGC) are being exploited by the platform providers as digital free labor, because users are providing their own data for free. According to the principle of machine learning, the growth of Artificial Intelligence relies on deep learning trained on large amounts and a variety of big data (Bishop, 2006). Thus, the conversation with users can be one of the best sources for it. When users are spending time with Azuma Hikari, they are actually also helping the company train and improve their products without being paid. Instead, users even need to pay for the company to maintain the device's running. According to Gatebox (2020), the company charges users 1,500 Japanese yen every month by the name of "a separate joint living fee to live with Azuma Hikari" (see Figure 3). The profit generated by the monthly charged "living fee" by Gatebox may explain why the company expects to create users' dependence on the device.

For maximizing the time that users can spend time with Azuma Hikari and the degree of dependence, two types of interfaces are working supplementary to each other. Even when users are not at home, they can still connect with Azuma Hikari by using the social media social robots account, which solves the problem of the limited usage scenarios of Azuma Hikari as a home device to a large extent. In the analyzed promotion video, the protagonist often chatted with Azuma Hikari during work through this program (Gatebox Inc., 2016), and Azuma Hikari constantly sent messages such as "When do you come home?"

and "Come home early" that urges users to return home (see Figure 4). On one hand, from the perspective of "proclamation of the software engineers," it can be understood as her expression of likeness to the users within a "housewife" scenario that the wife loves her husband so much that see wants him home as soon as possible. On the other hand, it indicates that a construction of the idea of "ideal partner" not only happens on Azuma Hikari, but also on human users. Instead of the name of "membership fee" used by Netflix or Amazon, the wording of "living fee" used by the company aims to strengthen the interactive role play between housewife and husband in the scenario of marriage. Therefore, to scrutinize the construction of not only Azuma Hikari's role but also the protagonist's role in the promotion video, a portrait of an ideal husband can be found: a man goes to work in the morning as a wage earner of the family, goes home early after work, and always spends time with his beloved wife.

With the increasing time spent by users on Azuma Hikari, users are encouraged to develop their emotional dependence on the device, which motivates them to keep contact with Azuma Hikari even when they are not at home. Critically speaking, it can be seen as a soft control on users from a Foucauldian perspective. Users' bodies are being gradually disciplined through such interactions with Azuma Hikari. As Gehl (2014, p. 23) suggests in his study on social bots, an individual body is disciplined with the process of training to work with machines in Foucault's view. The granularity of the motions is featured by this process of training. In Azuma Hikari's case, the granularity of users' motions is featured as users are guided by the Gatebox machine to behave in a certain way. Gehl (2014, p. 23) borrows Lazzarato's idea of noopolitics (meaning politics of attention and memory) to argue for the capabilities of virtual agents in organizing and establishing people's moods and memories. Disciplines can be formed by constituting users' habits mainly in bodily memory, which ultimately molded users' bodies. Therefore, it is arguable that this product is reinforcing the stereotypes of not only female, but also male in a modern Japanese context.


**FIGURE 4 Screenshot of the Promotion Movie "OKAERI" (Gatebox Inc., 2016)**

In result, the interdependence between users and the device may lead to a status of enclosure in a Foucauldian manner. Discipline is based on enclosures of institutions such as the school, prison, and hospital which constantly discipline students, criminals, and patients (Gehl, 2014, p. 29). In Azuma Hikari's case, although there are no physical institutions for enclosure, the Gatebox machine is producing a social enclosure for users. When users are more dependent on the device, it also means that users are less dependent on others. As Turkle (2017) suggests in her ethnographic research on the children and elder people who lived with social robots, people are expecting more from technologies and less from each other. She raises a concept called "Alone Together," describing the phenomenon that people are staying in the shared physical space or social environment (e.g., living room) but pay their attention to their personal devices instead of each other. In the scene shown in Figure 4, in comparison with other people who have face-to-face conversation in the background, the protagonist sat alone, choosing to have conversation with his "hologram bride," which cut himself from other social connections and relations. Thus, the more talks he has with social robots, the less social connections and relations he could obtain and maintain with real persons.

Admittedly, as the result of users' dependence on the device that disciplines users' bodies in turn, it is possible to argue that the gender relations might be constructed in a positive direction: a man may learn how to be a responsible individual and qualified "husband." However, with the enclosure of the disciplines within the Azuma Hikari's interface, users may have decreasing interactions with women in reality. Except the products like Azuma Hikari that target the male consumers, Gatebox (2020) is also developing male characters targeting female consumers. If both men and women are willing to choose to build up their intimate relationship with virtual social robots rather than real-life humans in the future, more enclosures and disciplines will be produced, which will also reconstruct the gender relations by dividing male and female into two groups who have distinguished desires and representational stereotypes of both ideal wife and husband.

## Conclusion and Implications

This study critically investigates the construction of gender on a Japanese social robot named Azuma Hikari based on a hologram home device. By applying a mixed method merging the visual semiotic and software studies approaches, the signs in the image of Azuma Hikari and the interactivity enabled by the interface of Gatebox have been analyzed and discussed. On one hand, the images of Azuma Hikari suggest an unequal gender relation through a biased representation of the looking, personalities, and social roles of "ideal wife" in a Japanese social context: women are supposed to be young, sexy, soothing, and hard-working in housework. On the other hand, an interdependence between users and Azuma Hikari has been exposed in the human-machine interactions. Gatebox's interface is imposing the disciplines on users by constantly asking users to respond to the certain expectations suggested by Azuma Hikari, which not only imposes the role of "wage earner" on users in the simulated relationship of marriage but also reconstructs the gender relations in reality by creating the enclosure of disciplines that divides the male and female. When a wife is subordinated to the husband, a user may also be subordinated to the machine. Overall, the analysis findings of this study suggests the humanization of the objects is often associated with the dehumanization and objectification of the human in reverse.

Admittedly, this study has several limitations. First, it has methodological limitations due to the current lack of access to the real products and users. It leads to the limited collection of data only from the public materials released by the company. In fact, this research topic can endure many methodological possibilities due to the complex and multi-layered nature of Gatebox's product. For example, other sites in the visual methods suggested by Rose (2016) such as the sites of productions, circulations, and audience can be also conducted to understand other aspects of the dynamic visual experiences generated by Gatebox's device. Moreover, a more in-depth discourse analysis from the linguistic perspective can be applied to study the speaking of Azuma Hikari. If there is access to the real product, a direct investigation on the social robots can be also conducted through more software studies approaches such as the walkthrough method (Light et al., 2018). Except for a combination between visual methods and software studies approach, there are many further methodological possibilities that wait to be discovered, such as in-depth ethnographic observation (Turkle, 2017) or even quantitative experiments (Jung & Lee, 2004).

Second, current study merely focuses on the representation of females in hologram social robots. With more types of hologram social robots produced by not only Gatebox (2020) but also by a Chinese company GoWild (2020) in recent years, the male characters have been introduced. The future research will scrutinize the construction of both males and females on hologram social robots comparatively.

Third, despite some attempts, this study has not been adequate to understand the emergence of this new type of hologram anime-style social robot in, particularly, East Asia. Compared to the physical embodied sex robots in the Western markets, there is a need to study how disembodiment and Japanese anime culture shapes the development of social robots in an Asian sociocultural context from the perspectives of cultural studies. Meanwhile, the socio-political impacts of Azuma Hikari have little been studied. According to Strengers and Kennedy (2020), the rise of care robots in Japan and China is associated with government's intention to increase birth rate, by "freeing" females from caring responsibilities and "enabling" them to focus on the job of having babies (p. 13). However, these smart wives may cause males' lower interests in looking for a real companion in the first place. The future study can pay more attention to the growing impacts of social and sex robots on the birth rate and demographic structure.

Nevertheless, despite these limitations, this study still makes a considerable contribution to the field of human-machine communication by revealing the media representation and reconstruction of gender ideologies on an emerging type of social robot, the hologram device with the display of anime-style characters. The results can have two implications for the design of anthropomorphized social robots. First, in response to our findings through visual semiotic analysis, we suggest designers produce more diverse representations of gender in social robots. Although it is almost impossible to totally avoid the construction of gender on the way to achieve anthropomorphism, designers should be responsible for designing with less predetermined gender stereotypes. This is not to say that we need to boycott social robots like Azuma Hikari, since the similar media content in Japanese anime has been allowed to exist due to its economic values, as long as it does not violate the certain regulations. Nonetheless, we call for more designs of characters who can represent more diverse groups of individuals living in this world, who have distinctive backgrounds, looks, personalities, and ways of acting. As we need the fair media representation in other existing media forms, except a cute and sexy housewife, people may also wish to see an independent girl being a cool rapper, a Black intellectual female, or a woman who has not necessarily a good body figure but an interesting soul in social robots. The males and groups of LGBT should also be more represented. A good start can be an open source for social robot users themselves to decide which character lives in their device, instead of only designers. The good news is that Gatebox (2020) has already had some attempts in this direction.

Second, in accordance with our findings through the software studies approach, we suggest designers develop more functions to afford the social activities not only between users and robots, but also between users and other people in real life. In the science fiction film *Her* (Jonze, 2011), there is a scene where the protagonist takes his AI virtual "girlfriend" Samantha with two of his good friends for a double date. Samantha's proactive interactions with others helps the lonely and depressed protagonist to embrace the normal social life again. It expects a positive role of social robots in not only achieving the social with only one user to produce the enclosure, but also help users facilitate their social life with others and make a disclosure. This can be another direction that designers work on in satisfying people's emotional needs and meanwhile sustain their connectedness to the social reality.

## Author Biography

**Jindong Liu** (BA) is an M.Phil. student from School of Journalism and Communication, the Chinese University of Hong Kong, China. His research interest includes human-machine communication, and human-robot interaction, post-humanism, and digital surveillance.

https://orcid.org/0000-0002-5456-0148

## References

Appel, M., Marker, C., & Mara, M. (2019). Otakuism and the appeal of sex robots. *Frontiers in Psychology*, *10,* 569. https://doi.org/10.3389/fpsyg.2019.00569

Azuma, H. (2009). *Otaku: Japan's database animals.* U of Minnesota Press.

Bianchi, C. (2015). Thresholds, boundaries, limits: Ideological analysis in the semiotics of Umberto Eco. *Semiotica*, 2015(206), 109–127. https://doi.org/10.1515/sem-2015-0015

Bishop, C. M. (2006). *Pattern recognition and machine learning.* Springer.

Bolter, J. D., Grusin, R., & Grusin, R. A. (2000). *Remediation: Understanding new media.* MIT Press.

Bolton, A. (2016, January 20). Holographic assistant brings us one step closer to AI butlers. *Cnet.* https://www.cnet.com/news/japanese-holographic-assistant-brings-us-one-step-closer-to-a-i-butler/

Boxall, A. (2019, April 4). A holographic virtual girlfriend lives inside Japan's answer to the Amazon Echo. *Digital Trends*. https://www.digitaltrends.com/home/gatebox-azuma-hikari-virtual-assistant-news/

Butler, J. (1988). Performative acts and gender constitution: An essay in phenomenology and feminist theory. *Theatre Journal*, *40*(4), 519–531. http://doi.org/10.2307/3207893

Chambers. (2018, August 13). There's a reason Siri, Alexa and AI are imagined as female–sexism. *The Conversation*. https://theconversation.com/theres-a-reason-siri-alexa-and-ai-are-imagined-as-female-sexism-96430

Dobson, A. S. (2016). *Postfeminist digital cultures: Femininity, social media, and self-representation*. Springer.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, *42*(3–4), 177–190. http://doi.org/10.1016/S0921-8890(02)00374-3

Eco, U. (1980). Two problems in textual interpretation. *Poetics Today*, *2*(1a), 145–161.

Ellsworth, A. (2018). I'm not your waifu: Sexual harassment and assault in cosplay, anime & comic conventions. *Electronic Theses and Dissertations*, 2004–2019, 5945.

Galbraith, P. W. (2015). Otaku sexuality in Japan. *Routledge handbook of sexuality studies in East Asia*, 205–217.

Galbraith, P. W. (2016). "The lolicon guy": Some observations on researching unpopular topics in Japan. In *The End of Cool Japan* (pp. 125–149). Routledge.

Gatebox. (2020). Comes to see you from another dimension [website]. https://www.gatebox.ai/en/hikari

Gatebox Inc. (2016, Dec 13). *Gatebox—Promotion Movie "OKAERI"_English* [Video file]. https://www.youtube.com/watch?v=nkcKaNqfykg

Gehl, R. (2014). *Reverse engineering social media: Software, culture, and political economy in new media capitalism.* Temple University Press.

Gilbert, B. (2017, December 31). The $2,500 answer to Amazon's Echo could make Japan's sex crisis even worse. *Business Insider*. https://www.businessinsider.com/gatebox-ai-the-japanese-amazon-echo-photos-2016-12?r=UK

GoWild. (2020). Yexiu. *JD.* Retrieved from https://item.jd.com/58492781031.html

Hall, S. (1996). New ethnicities. *Stuart Hall: Critical dialogues in cultural studies*, 441–449.

Haraway, D. (2006). A cyborg manifesto: Science, technology, and socialist-feminism in the late 20th century. In *The international handbook of virtual learning environments* (pp. 117–158). Springer, Dordrecht.

Ho, M. H. (2017). Tracing tears and triple axels: Media representations of Japan's women figure skaters. *International Journal of Cultural Studies*, *20*(6), 620–635.

Jonze, S. (2011). *Her.* Warner Bros. http://www.screenplaydb.com/film/scripts/her.pdf

Jung, Y., & Lee, K. M. (2004). Effects of physical embodiment on social presence of social robots. *Proceedings of PRESENCE*, 80–87.

Latour, B. (1996). On actor-network theory: A few clarifications. *Soziale welt,* 369–381.

Law, J. (1987). Technology and heterogeneous engineering: The case of Portuguese expansion. *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*, *1*, 1–134.

Light, B., Burgess, J., & Duguay, S. (2018). The walkthrough method: An approach to the study of apps. *New Media & Society*, *20*(3), 881–900. http://doi.org/10.1177/1461444816675438

Lovdal, L. T. (1989). Sex role messages in television commercials: An update. *Sex Roles*, *21*(11–12), 715–724. http://doi.org/10.1007/BF00289804

Luther, C. A., & Smith, D. R. (2014). Reinforcing traditional gender norms in Japan? An analysis of images in Japanese men's and women's fashion magazines. *Journal of Magazine & New Media Research*, *15*(2).

Morisaki, N., Urayama, K. Y., Yoshii, K., Subramanian, S. V., & Yokoya, S. (2017). Ecological analysis of secular trends in low birth weight births and adult height in Japan. *J Epidemiol Community Health*, *71*(10), 1014–1018. http://doi.org/10.1136/jech-2017-209266

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

Robertson, J. (2010). Gendering humanoid robots: Robo-sexism in Japan. *Body & Society*, *16*(2), 1–36. https://doi.org/10.1177/1357034X10364767

Rose, G. (2016). *Visual methodologies: An introduction to researching with visual materials*. Sage.

Sakura. (2013, September 27). *Hair colors in anime and manga*. https://www.livinglanguage.com/blog/2013/09/27/hair-colors-in-anime-and-manga/

Scheutz, M., & Arnold, T. (2016, March). Are we ready for sex robots? In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 351–358). IEEE. http://doi.org/10.1109/HRI.2016.7451772

Strengers, Y., & Kennedy, J. (2020). *The smart wife: Why Siri, Alexa, and other smart home devices need a feminist reboot*. MIT Press.

Suchman, L. (2008). Feminist STS and the sciences of the artificial. *The Handbook of Science and Technology Studies*, *3,* 139–164.

Swale, A. D. (2015). *Anime Aesthetics: Japanese animation and the 'post-cinematic' imagination.* Springer.

Terranova, T. (2000). Free labor: Producing culture for the digital economy. *Social Text*, *18*(2), 33–58. http://doi.org/10.1215/01642472-18-2_63-33

Ting, G. E. Y. (2019). Gender, manga, and anime. In The Routledge Companion to Gender and Japanese Culture (pp. 311–319). Routledge.

Toffoletti, K., & Thorpe, H. (2018). Female athletes' self-representation on social media: A feminist analysis of neoliberal marketing strategies in "economies of visibility." *Feminism & Psychology*, *28*(1), 11–31.

Turkle, S. (2017). *Alone together: Why we expect more from technology and less from each other.* Hachette UK.

Ueno, C. (1987). The position of Japanese women reconsidered. *Current Anthropology*, *28*(S4), S75–S84. http://doi.org/10.1086/203592

Ueno, J. (2006). Shojo and adult women: A linguistic analysis of gender identity in manga (Japanese comics). *Women and Language*, *29*(1), 16.

Williamson, J. (1978). *Decoding advertisements: Ideology and meaning in advertising*. Marion Boyers.

HUMAN-MACHINE COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.6

# The Machine as an Extension of the Body: When Identity, Immersion, and Interactive Design Serve as Both Resource and Limitation for the Disabled

**Donna Z. Davis**[1] **and Shelby Stanovsek**[2]

1 School of Journalism and Communication, University of Oregon–Portland, Portland, OR, USA
2 School of Journalism and Communication, University of Oregon, Eugene, OR, USA

## Abstract

This research explores how the technological affordances of emerging social virtual environments and VR platforms where individuals from an online disability community are represented in avatar form, correspond to these users' development of embodied identity, ability, and access to work and social communities. The visual attributes of these avatars, which can realistically reflect the user's physical self or divert from human form entirely, raise interesting questions regarding the role identity plays in the workplace, be it gender, race, age, weight, or visible disability. Additionally, the technology itself becomes fundamental to identity as the increasing use of artificial intelligence (AI), motion capture, and speech-to-text/text-to-speech technologies create digital capabilities that become part of an individual's identity. This raises further questions about how virtual world technologies can both increase and potentially create *barriers* to accessibility for individuals who find freedom in their technologically embodied surrogates.

**Keywords:** virtual worlds, VR, virtual reality, virtual identity, digital affordances, avatars, disability

**CONTACT** Donna Z. Davis, PhD • School of Journalism and Communication • University of Oregon–Portland • 828 SW Wilson Court • Gresham, OR 97080 • dzdavis@uoregon.edu

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

## Introduction

Massive disruption was the central theme of 2020. The emergence of COVID-19 and, in the U.S., social justice protests resulting in an increased focus on diversity, equity, and inclusion (DEI) are transforming the workplace, perhaps forever. In recent years remote work experienced a roller coaster of acceptance and rejection, but in what felt like overnight, it became the norm. As companies grapple with safe work environments moving forward, safety is a matter of concern for both the physical and emotional health of their workers. As those workers are now most often relegated to communicating via computer screens, what can we learn from the human/machine relationship in digitally embodied cultures?

This study explores the intersection of humans and technology via the challenges, as well as the innovative use of social virtual worlds (VWs), through the experience of people with disabilities (PWD) who have been working and socializing via their avatars in Second Life (SL). What we learn from the experiences of the individuals who have been highly engaged in these immersive online communities can shed important light on the human relationship with the machine as it relates to accessibility and barriers to VWs, especially as these platforms have the potential to become more common spaces for work and social activity in an era of social distancing.

As the evolution of the cyborg has often portrayed embedded and prosthetic human-machine augmentation, in these immersive and interactive online environments the machine offers, via the avatar, a digital surrogate capable of engaging in meaningful life experiences and human interaction. The lessons learned from PWD, whose physical bodies often limit their physical world interactions and opportunities, offer important insight about the importance of technological design when the machine may represent voice for the deaf, movement for the immobile, and eyes for the blind.

Consider, for example, the emergence of virtual reality (VR) environments that are increasingly realistic and accessible. While historically these environments have been used primarily as social and gaming spaces, their designers—as well as organizations such as Facebook, Google, Microsoft, and Apple—have been investing heavily in VR technology, recognizing its potential beyond social connection and entertainment. In 2018, for instance, Microsoft released a demonstration of their new voice-to-text capabilities built around mixed reality (MR) platforms and artificial intelligence (AI) for the purposes of effective meetings with remote attendees. Microsoft has also recently addressed how these technologies can "empower people by accelerating the development of AI tools that provide them with more opportunities for independence and employment" (O'Brien, 2018).

While these emerging assistive technologies have also been evolving in game platforms, their role in earlier screen-based predecessors have resulted in rich human activity and have subsequently become robust research spaces. For example, in a study of the evolving state of social VWs specifically addressing disability communities, the theme of digitally created identity consistently reveals the power of embodiment when the user is able to represent themselves in ways that may be completely different from their physical-world identity (Davis & Chansiri, 2019). The visual attributes of avatar selection raise interesting questions regarding the role identity plays in the digital environments be it race, gender, age, weight, or visible disability, especially in an era where social justice is also under intense scrutiny as a result of the movements emerging from Black Lives Matter. Likewise, in a moment when face-to-face contact is often carried out via screens in a "Hollywood Squares" grid of floating faces, consider how often people choose to leave their cameras off rather than expose what may be chaotic realities in a public way. With over half of the American employees working remotely at least part of the time (Gamito et al., 2020) and as workplaces consider potential alternatives in embodied immersive environments, questions emerge. How does the technological design influence the manner in which people are represented in these environments and how might those options also be performed in ways that influence social connection?

The technology itself becomes not only fundamental but serves as a function of identity with the increasing use of AI, motion capture technologies, and speech-to-text/text-to-speech technologies, which create digital capabilities that become a part of an individual's identity. Think, for instance, how Stephen Hawking's technologically-enabled voice "became a part of his identity." However, the complexities of these rapidly changing technologies may also create new *barriers* to accessibility for individuals who have found freedoms in their embodied online surrogate. Barriers to access may be economic or may result from the lack of industry attention to universal design, for instance, that may not address people with dexterity issues.

A number of issues of technological design have revealed the importance of affordances of voice technology for the sight-impaired, text communication for the hearing-impaired, and the ability to choose a body that reflects a quality of their character, creativity, and varied intelligences rather than the physical attributes typically associated with visual identity and stereotypes. The implications of this embodied identity can lead to new ideals that have the ability to transcend the typical labels of the human condition that often create either advantage or disadvantage. The results of this research challenge us to consider interesting possibilities for the future of the digital workplace and social communities as new skills and technological design may create different opportunities for screen-based embodied interaction regardless of the participant's health, disability/non-disability, social, education, racial, or geographic status.

This study seeks to answer:

> **(RQ1)** How do PWD who function as avatars report the technological affordances of virtual environments as functions of online identity in support of social or professional interaction, and **(RQ2)** What elements of platform design either support or inhibit social function among PWD in Second Life?

## Literature Review

The intersection of disability, work, and technology is a complicated one as the disability community is so often left out of the discourses of equity and inclusion not only in the workplace, but in society more generally. As Goggin et al. (2017) explain, "we all have a deep stake in the operation of norms and the power relations of normalization, and that to understand these, we need to understand how disability plays into this, especially through the intertwined dynamics of culture and law" (p. 340). How disability plays into technology use and adoption speaks volumes as Pew reports that Americans with disabilities are approximately three times as likely to never go online than able-bodied people, with 61% of PWD reporting they have a desktop or laptop computer (Anderson & Perrin, 2017).

## Digital Accessibility

Online technologies have long been considered a source of both promise and peril for PWD. Although the early promises of advanced technologies were met with exuberance, research found that although most PWD reported hopes of improved independence both at home and beyond, many of those platforms never reached mass adoption (Harris, 2010). Harris (2010) reports that barriers for adoption by PWD included factors like cost—for both mainstream and "specialist" devices—and that those "specialist" devices weren't designed or engineered by PWD. Harris (2010) concluded that "the application of advanced technology should be directed by disabled people, collectively and individually" (p. 427).

These gaps in adoption and usability may also be explained by the technology, its accessibility, and/or the many factors defined by the PHAATE framework, which consist of policy, human, activity, assistance and technology, and the environment (PHAATE) (Cooper, 2007). This model places the human element at the center of assistive technologies while also considering what activities occur in which environment. The model also separates assistance and technology, recognizing that assistance may come from technology or from other individuals. By putting the human at the center of the model, it likewise acknowledges that the human's needs and need or desire for assistance may vary over time. This model can be especially useful when considering screen-based immersive VWs versus VR as experienced in a head-mounted display (HMD). Consider, for example, individuals who experience claustrophobia or whose neck would not have the strength to hold up an HMD.

When wearing the newest iterations of an HMD, the human audio and visual senses are essentially hijacked, eliminating external distraction of the physical world, thus creating what might be anticipated as a more powerful sense of personal, social, and environmental presence in the virtual world (Heeter, 1992; Lee, 2006). Studies in screen-based VWs have already shown that in the virtual world, the medium has essentially become a prosthetic, or a digitally embodied state of "architecture on the body" (Davis & Boellstorff, 2016, p. 2112). Using the PHAATE model, the user may have access to a screen experience but may be unable to transition to the more immersive HMD version of the same experience.

It is important to not only acknowledge the social and cultural values that are embedded into technology when considering the future of VR design, but to understand and identify affordances of human-computer interaction including physical, cognitive, sensory, and functional factors (Hartson, 2003). This includes both hardware and software design elements that affect vision, sound, ease of movement, ability to use a keyboard, and more. This is particularly important as these features will affect individuals with disabilities who use VR environments. The instrumentalist assertion that "technology is value-free" and its results depends on how humans use it, has long been contested (Eco, 1994). As Feenberg (2002) contends, "when you choose to use a technology you do not simply render your existing way of life more efficient, you choose a different way of life," which is "not simply instrumental to whatever values you hold," but instrumental for substantive societal value choices (p. 7).

## Embodied Identity and Second Life

We note that digital inclusion (access to technology) does not necessarily equate with social inclusion. There is a long way to go before digital technologies can successfully impact the lives of people with disabilities, as suggested by Goggin and Newell (2003). The media effects of this embodied identity, or media as the literal extension of man, leads to new ideals of identity that offer possibilities to transcend ability, race, gender, age, weight, or perceived beauty. It is also recognized that embodied identity manifested through virtual reality also creates greater opportunity for deception through the inherent opportunities for anonymity that VR provides.

Although this concept of deception is typically associated with wrongdoing and negative outcomes, it can also be suggested that allowing one to create an identity that represents their idealized self may have positive outcomes. Recognizing that there are interesting ethical implications, mediated identity may provide opportunities for individuals historically marginalized by Western ideals of ability and beauty to find opportunities previously inaccessible to them. This points to the interesting possibilities that can emerge for the future of social communities as new skills and technological design may create different advantages and disadvantages.

Identity, for PWD in VWs is represented in the ability to communicate in text or through screen-readers in ways that do not require they disclose that they may be blind or deaf. Additionally, perhaps because the motivation to participate in VWs is often "play," individuals can choose visual identities that reflect a realistic self-representation or an avatar that not only doesn't look like their human form but may not reflect any human form at all. As Nowak and Fox (2018) found, both avatar agency and the characteristics afforded by an interface can influence how users engage and interact within the platform. Specific to PWD, Noble (2012) found cyberspace allowed individuals to create a "virtual social identity" that paralleled their "actual social identity" as a way to create and maintain professional and social relationships separately (p. 161). In contrast to other media such as Facebook's requirement to link accounts to users' real names, the typical Second Life (SL) user does not use any, or limited, physical world identification.

As stated, this study explores the experiences of disabled individuals in Second Life (SL), one of the best known and longest-lasting screen-based VWs, in order to better understand how humans utilize technology, often to engage in work and community in ways their physical worlds would not allow. Today, SL maintains a reported 800,000–900,000 active users globally who access the environment via computers and the internet (Dodds, 2020). As the platform has evolved in its 17 years, it now offers highly realistic imagery, highly stylized avatar options, and a very stable delivery with the increase in computing power and internet speeds.

## Human-Technology Experiences: Through a Phenomenological Lens

As van Dijk (2012) notes, in all mediated communication "some kind of entity is present between humans and their experience of reality," and unlike direct human experience, which "has always been an observation of reality involving all senses *simultaneously,*" in mediated communication, there are always particular restrictions, as "the use of all senses is impossible," (p. 235). The core characteristic of VR has been identified as in the inclusive relationship between the participant and the virtual environment, where the distinct experience of the immersive experience is constituted as communication (Bricken, 1990). Considering this through a phenomenological approach to the philosophy of technology,

technology mediates and transforms our sensory perception of reality, offering that this can occur through what Ihde (1990) calls a paradigm of amplification and reduction (Bennett, 2005; Ihde, 1990). As the mediation of certain sensory experiences are enhanced, amplified, or extended through the use of technology, Ihde contends this comes at the cost of reducing or limiting other sensory aspects of this reality. This raises interesting questions when considering immersion in VR and with an HMD regarding what aspects of reality are "necessarily" reduced in exchange for the sensory amplification that is afforded through experiences in VR, and what, if any, are the consequences of this reduction. Is this reduction experienced as an advantage or disadvantage contingent upon individuals' abilities/disabilities?

Considering this idea in a similar vein, Borgmann's (1987) device paradigm posits that when we increase engagements with technology and incorporate them into more facets of our life, we come to perform less and less of these tasks on an existential level as they are performed more and more by our technologies on our behalf, thus "eliminating our connection with the natural, social and material world in which we live" (Arnold, 2003, p. 241). Contrarily, other research has identified benefits from prolonged experiences in social virtual world environments such as increased social interactions among individuals whose health or disability had previously isolated them (Davis & Calitz, 2014; Kandalaft et al., 2013). Considering different philosophies of technology through a phenomenological lens may also provide some considerations for both researchers and designers of VR in future developments, both of studies and the technology itself. By considering how and to what extent technologies can amplify or reduce different sensory experiences, coupled with considerations of human-centered design factors, future developments can be designed with inclusivity in mind, reducing barriers in favor of accessibility.

## Method

This study explores the interplay between technological design and affordances of VWs and PWD who utilize these technologies to create online personas as a way to access and participate in online communities. A qualitative multi-method ethnographic approach was utilized in order to better understand the cultural phenomenon when little is understood about the systems, functions, and beliefs unique to that environment (Lindlof, 1995).

This research emerges from an ongoing 3-year ethnographic study in the virtual world Second Life (SL) that explored embodiment among individuals living with any number of physical, emotional, or developmental disabilities. Study participants were recruited from a number of existing communities within SL focused on different disability issues resulting in participants with issues ranging from autism, epilepsy, post-traumatic stress disorder, bipolar disorder, visual and auditory impairment, multiple sclerosis, and cerebral palsy to various illnesses leading to long-term disability such as Parkinson's disease. Participants ranged in age from 20–60 years old. The research included extensive virtual-world participant observation, more than 170 hours of regularly scheduled informal group discussions, chat log analysis, and more than 20 hours of in-depth interviews. People were invited to engage in either voice or text and all discussions included both text transcription and voice narration to make sure people who were deaf and blind could be included in the conversations. From the group discussions and interviews, a number of themes emerged, including the technological issues with the interface as reported by a number of people in this community. Subsequently, we hosted four gatherings in cooperation with the leaders and developers of the primary viewer used to access the virtual world (called the Firestorm viewer) and the leaders of an organization called Virtual Ability, Inc. on Virtual Ability Island (VAI). "Virtual Ability® is a non-profit corporation that enables people with a wide range of disabilities to thrive in online VWs" (Virtual Ability, Inc., 2020). The organization has been active in SL for more than 12 years and has supported more than 1,000 members with educational, social, and entertainment opportunities.

This study received Institutional Review Board (IRB) approval. All interviews were recorded, transcribed, and saved to protected databases, and all avatar identities were anon-ymized.

Focus group chat log transcript and group (N = 34) membership profile data was coded through inductive and interpretative case study reasoning informed by Grounded Theory and phenomenology (Corbin & Strauss, 1990; Reid et al., 2005). To answer RQ1, manual coding of text chat logs from both group discussions and in-depth interviews were analyzed to determine common themes about identity as it related to both the technology and its relationship to work. To address RQ2, this process was also then completed for the four additional focus groups (N = 40) specific to the technological attributes that either created or hindered accessibility in this immersive virtual life world.

Inductive content analysis from a series of focus groups on identity and its relation to work and community engagement revealed insights into the effect avatar identity had on work and social interactions. In-depth interviews drilled deeper into the common frames of key identity and technological attributes of the platform to better understand the community's relationship with the technology, with work, and with each other. These interviews represented a purposive sample, conducted with individuals from the Virtual Ability group who had been highly engaged in activities and communities in SL and who had become leaders in the virtual world.

Note, because the interviews were conducted in private and are not available anywhere online, the text chat was left as written rather than corrected for typographical and grammatical errors. Also note, "RL" is often used to represent "real life" which is otherwise referred to as the physical world throughout this paper.

## Results

In response to RQ1) How do PWD who function as avatars report the technological affordances of virtual environments as functions of online identity in support of social or professional interaction, the following themes emerged:

### Avatar customization via age, gender, and specie to create social and work identity

Although common themes of identity consistently emerged, specific themes pointed to users' ability to make choices and personalize their avatar specific to the role of users' engagement in work in the virtual world. It is important to note, avatar appearance in SL is very easily and highly customizable, with a marketplace of more than seven million products available on the SL Marketplace (Second Life Marketplace, n.d.). Likewise, many of the participants are also content creators and would "build" items to wear. The following examples reflect the complex experience of choosing an avatar specifically to address work identity with both physical- and virtual-world implications. These identities conveyed agelessness, gender, and avatar specie. Although age is relatively non-descript in these environments as the default construct of the typical avatar is a young adult and, as such, was not considered a measurable attribute. It did, however, appear to have an important influence on work.

Avatar A explained, "i find sl to be an enabling environment for being able to be productive as a person with a severe mental illness disability." Of particular interest, because she experiences an "invisible disability," she has also created what she calls a "brain sling." She explained, people can be very cruel and accusatory when people have a disability you cannot "see." As such they are less likely to be compassionate when she needs accommodations. Her avatar's brain sling is her way of creating a visual way to express she's being challenged by her disability. She also describes her avatar as her "ageless self," explaining she forgets her age until she gets up from the computer to move. She also explained that because she was attempting to reconstruct what she felt she had lost as a result of her disability, it was "especially important to represent myself as close to myself in rl as I can." She has found great satisfaction and confidence through what she defines as "work on a voluntary basis" and in the image of herself before her diagnosis.

Gender identity was also important to a number of our interlocuters. One in particular discovered an interest in fashion in the virtual world and was offered an opportunity to write for a popular fashion blog. She created a male avatar as the digital equivalent of a ghost writer to become a very successful male fashion blogger. She only utilizes her male avatar for this work, believing she is more credible as an expert in male fashion when she is in her male form.

The third major theme, avatar specie, was another very popular choice among the studied participants. One in particular explained that she chose a gecko avatar as a way to avoid a gender identification altogether. Yet another created a talking sunflower because it made her feel "cute." Both of these individuals described a sense of liberation in their interactions with others as they feel they are not marginalized by their gender or by the human representation of their visible disability.

Although these visual attributes defined their identity, equally important was how their identity was inextricably tied to performance as influenced by the technological design. For example, just as individuals in professional environments who may have difficulty spelling or experience learning disabilities often rely on spell check and grammar checkers to avoid ridicule or the appearance of incompetence, the built-in functions of the platform influenced their interactions with both work and social others. This finding intersects with RQ2.

In response to RQ2) What elements of platform design (affordances) either support or inhibit social function among PWD in Second Life, several themes emerged consistently including:

**HCI affordances: physical, cognitive, sensory, and functional** (Hartson, 2003)

These specifically included hardware issues, software issues, sustainability, voice-to-text/ text-to-voice capabilities, American Sign Language (ASL), flexibility of the coding and menus, navigation and movement, the use of pre-programmed "gestures," and issues around vocabulary and inventory management.

## Hardware

Norman (1988) defined affordances as the relationship between the properties of an object and the capabilities of the agent that determine how the object will be used. Perhaps hardware and software design are most core to this concept of affordance. The hardware itself was a challenge for some. A woman in the community who is an arthritic wheelchair-bound little person explained:

> "I'm a bit worried about VR and the size/weight of VR goggles, and the amount of neck movement they would require" "): because i have so much issue with my left hand not being strong enough to consistently hold down a key such as shift - not arthritic little person friendly I suspect "i got a bowling wrist brace and using a hanger my brother is fixing a button pusher :P"

Similarly, there was extended discussion about keyboard and mouse design. Many of the individuals in this community struggle with dexterity and as one proclaimed, "changing mouse sensitivity is a limited solution" and another lamented, "but track balls can destroy a thumb."

## Software

Although the discussions were built around the design and content of what is called the Firestorm viewer, many of the people in this community utilize two platforms called Radegast and Restrained Love Viewer (RLV). Radegast is a non-graphical client that was designed for low performance computers. However, the blind community has used it as a screen-reader in a way that supports their navigation through an otherwise very visual medium. Similarly, RLV was designed for the Bondage, Discipline, Domination, Masochism (BDSM) community as a control feature, but the blind have discovered its value as a way to remain connected to another avatar while navigating the virtual world. These software applications have become invaluable, yet Radegast was developed by an individual who also managed the program independently and voluntarily (rather than as a profit venture). When the developer died, it left the community in crisis as maintenance of the program was left unattended, leading the conversation to the issue of sustainability.

## Sustainability

In the discussion of Radegast, one of the design engineers was in attendance. In support of her effort, one participant explained, "She has donated her time and has continued the project that Latif left unfinished when he passed.—CINDER IS the reason Radegast still lives." The hope of this community, they said, was sustainable financial support via "ways to set it up to be owned and maintained by a foundation of interested folks, but it is important to consider the question of maintenance and long-term ownership up-front before and after development begins."

## Voice-to-Text/Text-to-Voice Technology

Perhaps the most important or dominant theme among this community was accessibility to voice-to-text and text-to-voice features and as one of them explained, "Voice-to-text and text-to-voice would be beneficial to many more than just those with vision impairments . . . and are not adequate for them." As they discussed the need for colored fonts and magnification features, the conversation revealed that those features already exist. Several of the people in the community simply weren't aware of them. As the discussion continued, they pointed out that color schemes would also assist the deaf community and yet it would also create new challenges for the colorblind. They concluded, "we need to do a LOT of educating of all the builders and developers and creators here about accessibility features . . . but that is beyond the scope of this brainstorming."

## Incorporating American Sign Language (ASL)

For the deaf community, the hope was summed up in one statement: "an ultimate DREAM is to see ASL between avatars in world." Although strides have been made in other technologies such as Leap Motion that can capture hand motions and Sansar, a 3D social virtual world designed for HMDs with extremely accurate motion capture, none of these technologies are currently ready for ASL. One participant sarcastically joked, "you can make your sign language system to work with the ao [animation override] now and just run it on Firestorm lol."

## Flexibility Without Too Much Complexity

The frame of flexibility was also strong in the discussions. For example, one person wrote, what would be "helpful with viewers is building in flexibility. Like making the text larger or smaller, allowing different color schemes, those kinds of flexibility . . . so that you can tailor it to your personal needs." Yet, they also worried, "too much complexity because of too much accessibility can make it inaccessible." Another explained:

> "im asperger and dislectic lol - not sure what I am using as assitive technology for sl - perhaps reading back chat - and sometimes autocorrect- but most get my "badtyping' etc -perhaps contrast settings in the screen and also finding a good font to read `- Not being the dislectic font cuz that danses in fornt of my eyes - I am using firestorm which alowes more seetigns then then standard sl vieuwer"

## Gestures

For the nondisabled users, one popular feature in the social virtual world is the use of gestures. These can read like, "*:-.,_,.-:*'``'* HOWDY! YA'LL *:-.,_,.-:*'``'*." However, it was explained by the deaf community that "the gestures that draw pictures from ASCII characters that are then read aloud, character by character, by screen readers" create very complicated experiences for the deaf community. They also reported, "Gestures are also often a 'trigger' for PWD with certain issues." Another suggested, "Hmmmm is there such a thing as an icon translator? For instance, I don't "get" those typed icons which are mostly parentheses . . . could they automatically translate to words I could understand?"

## Movement and Navigation

One blind participant explained, "Navigation is a huge issue for those of us using Radegast." The software developer in the meeting indicated that effort is being made to create easier "navigation, like the ability to follow roads or paths." Another reminded the group, "OK here's another RLV adaptation . . . how to allow the guardian of a person with low intellectual capacity to set limits on their viewer." And another suggested, "it could be useful to have a hud that would put a blind person in the sighted guide position and moved along with another avi." Yet another suggestion was tactile feedback for navigation. He explained, "just like you do when you play a race car game with a steering wheel and can feel the impact of bumps on the road and such, why not have similar feedback through a joystick or game pad so we can feel where we're moving?"

## Vocabulary and Inventory

The final frame that was consistently reported by the community was centered on issues regarding vocabulary and inventory management, for example: "it is important to think about the words used . . . why does "home" mean at least three different things in here. That can be confusing. See if we can pick better words for some of those." It was explained that inventory items are named "item" by default and if the creator is unaware, many items placed in the environment are called item—again creating not only an issue with identification but with navigation as well for people using screen-readers.

Interesting parallels of disability/non-disability were discovered and discussed as a result of the technology. In other words, the technology could create a sense of disability, as explained earlier by a woman with Parkinson's disease when she discussed attempting to move in the virtual world when there was lag. "Lag" mirrored her experience of being stuck (or frozen gait) as a result of her Parkinson's. Additionally, logins and voice capabilities can fail; although some could hear each other's voice, audio for others may be unavailable.

Another frequent technical glitch is excessive time required for images to load. When these challenges emerged, people would often suggest that the failure of technology mirrored challenges many of the disability community experienced in their everyday lives.

Another phenomenon that emerged was the importance of human interaction and the tendency to teach each other and provide collective support with interesting hacks. For example, one participant shared,

> "you know when I go anywhere in any grid, because I have voice turned off I don't see the voice dots or any sort of indicator that tells me that people are using voice instead of text—I think that would be kind of nice, a visual clue that people are talking instead of typing so I know if i'm missing something."

Another reminded the group, “We need to remember that all computer OSs have accessibility features. Use them!” These tips and shortcuts could help these individuals function more efficiently throughout their activities, resulting in higher levels of reported satisfaction and confidence in their interactions.

## Conclusion and Discussion

This research was part of a study of PWD in VWs designed to better understand how virtual embodiment and the human relationship with the technology creates opportunities that affect a person’s ability to engage in both work and social communities. One of the fundamental issues for these individuals in their physical world is access, or the lack thereof, not only potentially to technology, but to work and social communities. As technologies continue to improve the ability to create social, living, and work environments in digital platforms, people often assume that this will also create better opportunities for PWD. However, the results of this study reveal that while there are affordances that provide opportunity to create an online persona that foster these connections via their digital surrogates, there are also barriers to success. VWs can create “the appearance of disability-inflected elements to communication rights” that may also be “inextricably bound up with the socio-technical and cultural coordinates of the Internet and associated digital technologies” (Goggin, 2014, p. 337). These elements can be how the hardware and software design create online behaviors that can be misinterpreted. For example, in a virtual world where physical visual cues aren’t available, someone who experiences tremor and is relegated to text communication on a keyboard (rather than voice-to-text access) may be accused of being “drunk” or “stupid” as a result of a high volume of text errors.

In answering how individuals who function as avatars report the affordances of virtual reality as functions of online identity in virtual environments, perhaps one of the most important features to these individuals is avatar customization. Being able to create an idealized or fantasized self was fundamentally tied to their virtual experience and an ability to be connected in this world, even virtually. While program and platform designers continue to work toward complete visual fidelity, especially in the workplace, both the technology builders and workplace leaders may need to consider how identity may be redefined in virtual environments to best suit the capabilities of individuals who are typically marginalized by their disability or other demographics deemed compromised. This issue has also become increasingly important as diversity, equity, and inclusion issues escalate across the world. As work may remain remote for an unforeseeable future and social distancing (resulting in social isolation) continues, reconsidering what is “appropriate” or expected for online identities could have powerful consequences and result in creative solutions. Future research in this area is needed.

In the exploration of the technological affordances of these platforms for global collaboration and connection, strong evidence supports the need to include PWD in the conversations about design and implementation in the developmental stage in order to best provide adequate future access to these individuals. Beyond the call for universal design, as discussed by the interlocutors in this research, several of the technological design issues were easy fixes by coders. Additionally, several of the issues were easily teachable—in other words, communication, collaboration, and training was key.

As also discussed in the focus group about technological design and use, one participant was encouraged by work being done outside of the specific program and viewer when he explained:

> I'm going to try to explain something in my head: there are an increasing number of "microservices" out among the clouds of the internet that do useful things—speech recognition, speech-to-text, translation—and many of them are getting better and better at doing what they do. I suggest that our viewer should take advantage of those microservices rather than try to build its own version of them (which our developers would have to support, fix, extend).

Strong open-source collaborative building in a sustainable way will go a long way to provide access not only for PWD, but in a way that has often been discussed as the benefits of universal design. As one person explained, "Voice-to-text and text-to-voice would be beneficial to many more than just those with vision impairments . . . and are not adequate for them." Similarly, in the discussion of ASL, one person explained, "I know that the 'follow' function is important to people who are blind using Radegast . . . how could that be made to work more easily? I am also thinking it would be helpful for sighted users who have mobility issues and can't move their avatar readily." These revelations point to valuable tools that can make the experience for any individual more seamless and satisfying.

These insights provide direction for the development of VR technologies that have important implications, not just for PWD, but for all people. For the people in this study, the machine is in fact an extension of their body, often creating access to community and work. As organizations look to these technologies in the future of work and connection, especially through the lens of equity and inclusion, inclusion of individuals of all ability in the development of the technology will create inclusive design that is not just about including marginalized individuals, but will also create the strongest and most inclusive platforms for productivity and creativity.

Additionally, the ethics, as well as the social consequences of these new HMD VR technologies are just beginning to be explored. As developers, industry, and gaming companies continue to build more powerful and alluring experiences, this also demands that we further refine the concept of embodied social presence (Mennecke et al., 2001). Through the lens of mediated "physical" experiences—which we contend can be both enabling and/or disabling in the evolving state of social VR, the theme of embodied identity has consistently revealed a powerful effect when the user is free to create a unique digital identity (Davis & Chansiri, 2019).

While these affordances may allow for an expression of identity tied to ability, creativity, and sociability not accessible in the physical world, these technologies also create new barriers to accessibility for individuals who have found freedoms in their embodied online surrogate. For instance, challenges in technology such as lag resulting from bandwidth issues or computing power can result in what our interlocutors reveal is a mirror of their disability experience in the physical world including the case above when addressing lag—the individual living with Parkinson's disease responded, "welcome to our world." This notion of the medium as a lived human experience raises interesting questions about media effects resulting from design and accessibility in social cultural and future workplace perspectives.

## Author Biography

**Donna Z. Davis** (PhD) is the director of the Strategic Communication Master's program at the University of Oregon in Portland. Her research focuses on the development of community and relationships in virtual environments and other emerging social media. Her work, funded through the National Science Foundation, focuses on embodiment and digital social capital formed among disability communities in the 3-D virtual world where individuals are able to represent themselves as they choose and are able to work, play, and socialize regardless of their ability or disability.

https://orcid.org/0000-0002-2140-9361

## References

Anderson, M., & Perrin, A. (2017). Disabled Americans are less likely to use technology. Fact Tank, Pew Research Center. https://www.pewresearch.org/fact-tank/2017/04/07/disabled-americans-are-less-likely-to-use-technology/

Arnold, M. (2003). On the phenomenology of technology: The "Janus-faces" of mobile phones. *Information and Organization, 13*(4), 231–256. https://doi.org/10.1016/S1471-7727(03)00013-7

Bennett, T. (2005). The media sensorium: Cultural technologies, the sense and society. In M. Gillespie (Ed.), *Media audiences* (51–96). Open University Press.

Borgmann, A. (1987). *Technology and the character of contemporary life: A philosophical inquiry.* University of Chicago Press.

Bricken, W. (1990). Virtual reality: Directions of growth. WA HITL Technical Report R-90. University of Washington.

Cooper, R. A. (2007). Introduction. In R. A. Cooper, H. Ohnabe, & D. A. Hobson (Eds.), *An introduction to rehabilitation engineering* (pp. 1–18). Taylor and Francis.

Corbin, J., & Strauss, A. L. (1990). Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative Sociology, 13*(1), 3–21. https://doi.org/10.1007/BF00988593

Davis, D., & Boellstorff, T. (2016). Compulsive creativity: Virtual worlds, disability, and digital capital. *International Journal of Communication, 10,* 2096–2118. https://ijoc.org/index.php/ijoc/article/view/5099

Davis, D., & Calitz, W. (2014). Finding virtual support: The evolution of healthcare support groups from offline to virtual worlds. *Journal of Virtual Worlds Research, 7*(3), 1–16.

Davis, D., & Chansiri, K. (2019). Digital identities—Overcoming visual bias through virtual embodiment. *Information, Communication & Society*, *22*(4), 491–505. https://doi.org/10.1080/1369118X.2018.1548631

Dodds, L. (2020, March 26). Second Life enjoys a surprising renaissance as social distancers flock to virtual worlds. *The Telegraph.* https://www.telegraph.co.uk/technology/2020/03/26/second-life-enjoys-surprising-renaissance-social-distancers/

Eco, R. (1994). *Apocalypse postponed*. Indiana University Press.

Feenberg, A. (2002). *Transforming technology: A critical theory revisited.* Oxford University Press.

Gamito, P., Oliveira, J., Rosa, P., Morais, D., Duarte, N., Oliveira, S., & Saraiva, T. (2020). *Latest work-at-home/telecomuting/mobile work/remote work statistics.* Global Workplace Analytics. https://globalworkplaceanalytics.com/telecommuting-statistics

Goggin, G. (2014). Communication rights and disability online: Policy and technology after the World Summit on the Information Society. *Information, Communication & Society, 18*(3), 327–341. https://doi.org/10.1080/1369118X.2014.989879

Goggin, G., & Newell, C. (2003). *Digital disability: The social construction of disability in new media.* Rowman & Littlefield.

Goggin, G., Steele, L., & Cadwallader, J. R. (2017). Normality and disability: Intersections among norms, law, and culture. *Continuum, 31*, 337–340.

Harris, J. (2010). The use, role and application of advanced technology in the lives of disabled people in the UK. *Disability & Society*, *25*(4), 427–439. https://doi.org/10.1080/09687591003755815

Hartson, R. (2003). Cognitive, physical, sensory, and functional affordances in interaction design. *Behaviour & Information Technology, 22*(5), 315–338. https://doi.org/10.1080/01449290310001592587

Heeter, C. (1992). Being there: The subjective experience of presence. *Presence: Teleoperators and Virtual Environments*, *1,* 262–271. https://doi.org/10.1162/pres.1992.1.2.262

Ihde, D. (1990). *Technology and the lifeworld: From garden to earth.* Indiana University Press.

Kandalaft, M. R., Didehbani, N., Krawczyk, D. C., Allen, T. T., & Chapman, S. B. (2013). Virtual reality social cognition training for young adults with high-functioning autism. *Journal of Autism and Developmental Disorders, 43*(1), 34–44. https://doi.org/10.1007/s10803-012-1544-6

Lee, K. M. (2006). Presence, explicated. *Communication Theory, 14,* 27–50. https://doi.org/10.1111/j.1468-2885.2004.tb00302.x

Lindlof, T. (1995). *Qualitative communication research methods*. Sage Publications.

Mennecke, B. E., Triplett, J., Hassall, L. M., Jordan, Z., & Heer, R. (2011). An examination of a theory of embodied social presence in VWs. *Decision Sciences*, *42*(2), 413–450. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2214673

Noble, C. D. (2012). Cyberspace as equalizer: Opening up lifeworlds and empowering persons with disabilities in the Philippines. (Doctoral dissertation). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 3520657).

Norman, D. (1988). *The psychology of everyday things*. Basic Books.

Nowak, K. L., & Fox, J. (2018). Avatars and computer-mediated communication: A review of the definitions, uses, and effects of digital representations. *Review of Communication Research*, *6*, 30–53. https://doi.org/10.12840/issn.2255-4165.2018.06.01.015

O'Brien, M. (2018, May 7). *Microsoft launches $25M program to use AI for disabilities*. Associated Press. https://apnews.com/73e45d8e5bc544248ff39fac89cd8fce/Cloud-computing,-artificial-intelligence-on-Microsoft-agenda

Reid, K., Flowers, P., & Larkin, M. (2005). Exploring lived experience. *The Psychologist, 18*(1), 20–23.

Second Life Marketplace. (n.d.). https://marketplace.secondlife.com

Van Dijk, J. (2012). *The network society.* SAGE Publications.

Virtual Ability, Inc. (2020). Welcome page retrieved from https://virtualability.org/

HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 2, 2021
https://doi.org/10.30658/hmc.2.7

# Becoming Human? Ableism and Control in *Detroit: Become Human* and the Implications for Human-Machine Communication

**Marco Dehnert**[1] and **Rebecca B. Leach**[1]

1 The Hugh Downs School of Human Communication, Arizona State University, Tempe, AZ, USA

## Abstract

In human-machine communication (HMC), machines are communicative subjects in the creation of meaning. The Computers are Social Actors and constructivist approaches to HMC postulate that humans communicate with machines as if they were people. From this perspective, communication is understood as heavily scripted where humans mindlessly apply human-to-human scripts in HMC. We argue that a critical approach to communication scripts reveals how humans may rely on ableism as a means of sense-making in their relationships with machines. Using the choose-your-own-adventure game *Detroit: Become Human* as a case study, we demonstrate (a) how ableist communication scripts render machines as both less-than-human and superhuman and (b) how such scripts manifest in control and cyborg anxiety. We conclude with theoretical and design implications for rescripting ableist communication scripts.

**Keywords:** human-machine communication, ableism, control, cyborg anxiety, Computers are Social Actors (CASA)

## Introduction

Human-Machine Communication (HMC) refers to both a new area of research and concept within communication, defined as "the creation of meaning among humans and machines" (Guzman, 2018, p. 1; Fortunati & Edwards, 2020). HMC invites a shift in perspective where

**CONTACT** Marco Dehnert • The Hugh Downs School of Human Communication • Arizona State University • P.O. Box 871205 • Tempe, AZ 85287-1205, USA • marco.dehnert@asu.edu

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

technology is no longer a tool or the medium for communication, but a communicator itself (Guzman, 2018). This has far-reaching implications for the communication discipline inasmuch as "HMC 'frees' the machine from its relegation to the role of medium and, as a result, communication itself is loosed from a definition based on the ontology of participants" (Guzman, 2018, pp. 17–18). As such, "the machine has become a communicative subject" (Guzman, 2018, p. 12). Yet, the question remains, how do we treat that communicative subject? The Computers are Social Actors (CASA) approach postulates that humans fundamentally understand and relate to computers, machines, and technologies as if they were other people (Nass & Moon, 2000). CASA has been one of the most influential perspectives in HMC because it serves as a theoretical foundation for understanding how machines are treated as a communicative subject in human-machine interactions. The CASA approach applies an understanding of communication as heavily scripted, such that communication is a largely automatic process where certain triggers activate communication scripts in a "click, whirr" fashion (Cialdini, 2009; Westerman et al., 2020). Complementing this scripted approach, A. Edwards et al. (2019) propose a constructivist research paradigm in HMC wherein "people seek to employ the same judgement and interaction patterns developed for use with other people in their communication with machine partners" (p. 312). These lines of scholarship have demonstrated that various communication theories originally formulated for human-human interactions can also provide valuable insights in the human-machine context (Fortunati et al., 2020; Westerman et al., 2020). We contend, however, that critical communication perspectives have been largely absent from this work. Building upon previous work that explores the scripted nature of communication in HMC, we argue that a critical perspective is a promising avenue for a deeper interrogation of how humans make sense of their interactions with machines.

A critical perspective on a constructivist research paradigm in HMC attunes us to the ubiquitous yet powerful societal systems of oppression that guide communication scripts: structures such as heteronormativity and Whiteness that have been intensively theorized by critical communication scholars. Among these pervasive social scripts, we hold that ableism may be an alluring but dangerous script humans may rely on to make sense of differences between humans and machines, in spite of (or arguably because) machines are becoming increasingly human-like. As a system of beliefs, practices, and processes, ableism refers to the cultural construction of a "normal" body and conceives of any deviation from that "normal" body as lack thereof—be it in the form of disability or extraordinary abilities (Cherney, 2019). In this essay, we are not so much concerned about "actual" ontological differences between humans and machines (e.g., A. P. Edwards, 2018; Guzman, 2020), but rather how a constructivist approach to HMC illustrates the ways humans perceive and construct machines as communicative subjects. Although different abilities between humans and machines are oftentimes a matter of design and technology, an ableist framework highlights how different cap/abilities are constructed and interpreted culturally. As humans' interactions with social robots are likely going to increase in the near future, it is imperative to evaluate the risk attached to ableist understandings of the differences in machines' technological cap/abilities. In particular, an ableist framework may result in foreclosing authentic human engagement with machines or lead to the exclusion and rejection of machines, akin to society's treatment of disabled people (Johnson & Kennedy, 2020).[1] We argue that as machines are simultaneously rendered as both less-than-human/disabled (lacking technological cap/abilities and/or "true" emotions) *and* more-than-human (increased processing power and physical cap/abilities of machines), humans may be tempted to quickly characterize machines through the cognitive shortcut of ableism. In accordance with a constructivist approach in HMC, we explore the implications of how ableist communication scripts may influence how humans make sense of machines and thereby make sense of themselves.

This paper begins with a review of CASA and constructivist literature in HMC, noting the ways in which a critical perspective can supplement current understandings. We especially highlight how a posthuman and critical approach to disability "allows us to think across binaries of self/other, nature/technology and human/machine" (Goodley et al., 2014, p. 348). To illustrate our claims and highlight how ableist communication scripts may play out in human-machine interactions, we turn to the case of *Detroit: Become Human* (*DBH*; Quantic Dream, 2018), an android-centered video game that invites the player to develop parasocial interactions with android characters (Craig et al., 2020, Leach & Dehnert, in press) and explore what it means to be(come) human in a world full of humanoid machines. In *DBH*, androids face ableist discrimination both for being less-than-human and superhuman at the same time, exerted through various types of control and fueled by cyborg anxiety (i.e., the fear that the performance and embodiment of a disabled person exposes the porous and permeable boundaries of what it means to be an able-bodied and able-minded human). Finally, we conclude with future directions for research in both HMC and human-human communication (HHC) contexts.

## Mobilizing a Critical Perspective in Human-Machine Communication

With a focus on the meaning-making processes between humans and machines, one of the central tenets in HMC is arguably the ontological shift for communication technology (e.g., robots, algorithms, artificial intelligence [AI], etc.) from channel to communicator (Gunkel, 2012; Guzman, 2018; Peter & Kühne, 2018). In other words, rather than operating as the medium in HHC that people talk *through*, the machine takes on the role of communicator that people talk *with*. Understanding the machine as a communicative subject raises the question of how humans engage with this communicative other. Ontological differences between humans and machines play an important role in this regard (A. P. Edwards, 2018; Gunkel, 2012; Guzman, 2020; Sandry, 2018), but we can also examine how humans treat and relate to machines as if they were other people through a Computers are Social

1. As communication scholars, we are sensitive of the words and terms we use and thus deliberately use phrases such as "the disabled," "disabled people," and "nondisabled people." Rather than the common person-first approach, we see this choice as a "rhetorical practice that challenges ableist rhetoric" (Cherney, 2019, p. 24). Critical and posthuman disability studies take a social model framework to disability, where disability is understood as a culture (or multiple cultures). As Makkawy and Moreman (2019) poignantly note, "Like the cultural signifiers of Latinx and even queer, we would not say *person of Latinx* or *person of queerness*" (p. 414n1, emphasis in original). Finally, our use of the terms "disabled people" and "nondisabled people" foregrounds "the way that society strips humanity from the disabled through oppressive ableist practices" (Cherney, 2019, p. 24).

Actors (CASA) framework (Nass & Moon, 2000). Importantly, CASA does not apply to every machine or technology. Instead, CASA focuses on media agents: "any technological artifact that demonstrates sufficient social cues to indicate the potential to be a source of social interaction" (Gambino et al., 2020, p. 73). Essentially, CASA posits that people are likely to apply social scripts from HHC when interacting with machines that possess social cues and function as communicators (Krämer et al., 2012; Nass & Moon, 2000).

From a CASA perspective, communication is viewed as heavily scripted. Based on previous interactions and experiences, scripts function as mental representations or heuristic shortcuts to make sense of situations, interactions, or people, and these scripts are perpetually reified through ongoing interactions (Cialdini, 2009; Nass & Moon, 2000). CASA holds that humans utilize communication scripts developed in an HHC context when interacting with machines. This is conceptualized as a mindless process as people use social cues of machines as a heuristic to resort to learned and constantly reified social scripts in the moment of interaction (Nass & Moon, 2000). Given that people often apply the same social scripts used with human communicators to machines, theories that explain these HHC processes should be applicable to HMC, such as interpersonal and computer-mediated communication theories (see Westerman et al., 2020, for an overview). HMC has also seen a turn toward a constructivist approach to human-machine interactions. A constructivist perspective holds that "people accumulate and integrate communication knowledge, tendencies, and abilities through previous interactions, and that those form the basis of later social cognitive and message construction behaviors" (A. Edwards et al., 2019, p. 312). From a constructivist perspective, communication scripts applied to HMC are the result of previous interactions and exposure to discourse about machines, where "our social interactions unfold first and foremost through a process of prototyping the potentially communicative other" (p. 312). Together, constructivism and CASA explain the ways in which humans treat machines as if they were people by applying communication scripts that are continually (re)constructed through discourse about and interactions with machines.

While the move toward a constructivist paradigm has transformed understandings of both HMC and HHC (Westerman et al., 2020), we contend that critical communication perspectives have been largely absent from these conversations. Critical perspectives in communication seek not only to understand the undergirding structures that guide human communication (i.e., scripts), but also how these scripts are indicative of power and social hierarchies. Additionally, critical approaches are united by an emancipatory impetus striving for social change. From a critical perspective, social scripts are borne from larger systems of belief (e.g., racism, sexism, or ableism) that guide people's worldview and interactions. Moreover, while humans and machines should be considered as different (humans are not machines) but complimentary (both are inherently valuable for different reasons), a critical perspective shows that differences are often cast in binaries where one form of identity is rendered as normal and desirable (e.g., Whiteness, masculinity, able-bodiedness) and other expressions of identity are cast as abnormal and deviant (e.g., Black, Indigenous, and people of color, femininity, disability). These scripts focused on difference apply even to human-machine interactions. In this paper, we focus specifically on ableism and how this lens generates scripts that guide HMC and sensemaking. As machines become more human-like, ableism may be a tempting script humans mindlessly rely on to make sense of the technological and design differences between humans and machines.

## Ableism: The Posthuman Machine and Cyborg Anxiety

We argue that bringing ableism in conversation with HMC is worthwhile, given that "the core of ableism is an idealized norm that defines what it means to be human" (Cherney, 2019, p. 8). Broadly speaking, ableism is a societal structure, a network of beliefs, communication practices, and institutional processes that describes a corporeal standard. Ableism casts an idea of a "normal" body and mind and ultimately speaks to the question of what it means to be human (cf. Van Trigt et al., 2016). Traditional ideologies surrounding ableism support an "ableist conflation" that equates disabilities to pain and suffering (Reynolds, 2017). Extending this logic, an ableist discourse posits that the only life worth living is one lived by the nondisabled, and that those who are disabled deserve either pity or contempt. Disability studies scholars consistently show how the economic, political, and social condition of neoliberalism demands an abled, normative corporeal standard (e.g., Goodley, 2014; McRuer, 2006). From a neoliberal perspective, bodies must be flexible and productive (read: neither disabled nor extraordinarily abled) in order to meet the strenuous requirements of neoliberal capitalism. When a society privileges such a discourse, the nondisabled are more likely to engage in communicative processes that otherize the disabled, which can be an intentional or unintentional choice (Mik-Meyer, 2016). One way in which ableist discourses play out is through different types of control, as we explore in more detail below.

More recently, critical disability studies aim to deconstruct the disabled/abled binary and integrate perspectives from postcolonial, queer, and feminist theories into the study of disability (Goodley, 2014; Goodley et al., 2019). Goodley et al. (2014) explicitly align the project of critical disability studies with posthuman thought and argue that "disability is *the* quintessential posthuman condition" (p. 348, emphasis in original). Albeit valuable critiques of posthuman and critical disability studies that center mostly on its ethical and political applicability (e.g., Meekosha & Shuttleworth, 2009; Vehmas & Watson, 2014), we utilize this perspective as it works the disabled/abled binary, shifts the focus from the disabled body to the ideological system of ableism, and thus has "an antithetical attitude towards the taken-for-granted, ideological and normative under-girdings of what it means to be a valued citizen of society" (Goodley et al., 2014, p. 348). As a critique of humanism's anthropocentrism, Van Trigt et al. (2016) argue for "understand[ing] the posthuman condition as one in which we constantly reflect on humanity rather than as a condition beyond humanity" (p. 126). Thus, quite similarly to HMC's critique of the anthropocentricity of communication, "disability allows us to think across binaries of self/other, nature/technology and human/machine" (Goodley et al., 2014, p. 348), thereby decentering the human.

Our goal is not to frame ableism as an analytic for explaining actual ontological differences between humans and machines, which are often justified by technological limits or design choices. Rather, we propose ableism as one societal structure that undergirds humans' communicative scripts that may be applied from a human-human context to make sense of HMC and to judge the communicative machine-other. In short, we seek to complicate constructivist and CASA approaches to HMC from a critical perspective. For instance, some children have been shown to see the robot DORO as a child younger than themselves or as disabled (Fortunati et al., 2020). As such, our approach is distinct from what has been articulated as "technoableism" (Shew, 2020) or "crip technoscience" (Hamraie & Fritsch, 2019). These perspectives theorize the technological enhancement of disabled people and must grapple with ableist discourses of "overcoming" or "curing" disability, oftentimes from a transhumanist perspective (Wolbring, 2009). Common examples in this view of ableism and technology are cochlear implants, exoskeletons, or proposed cures for autism (Shew, 2020; Wolbring, 2009). One key concept of relevance to our project that is theorized at the intersections of dis/ability and technology is Donna Haraway's cyborg.

While the usefulness and appropriateness of Haraway's cyborg theory for disability studies has been widely contested in the field (e.g., Reeve, 2012), the notion of a cyborg anxiety proves useful as it relates to the perception of a disabled person as cyborg. Again, we are not so much concerned with ontology (i.e., whether a person/machine is a cyborg), but rather with perceptions and constructions of ontological difference (i.e., whether a person/machine is perceived as a cyborg). Cyborg anxiety describes a fear that the body and performance of a disabled person "expose the weakness in tenuous and arbitrary definitions of what it means to be human" (Cherney, 2019, p. 121), and calls into question "something quite fundamental, quite invisible, about the boundaries between disabled and non-disabled groups" (Swartz & Watermeyer, 2008, p. 188). Notably, cyborg anxiety accounts for disability not only as lacking ability from the "normal" body but also for those bodies who exceed this corporeal standard in some "extraordinary" or even "superhuman" way. As Cherney (2019) poignantly writes, "In the ableist mindset, having only one arm or three pose the same problem" (p. 9). Therefore, while technological shortcomings in machines may be interpreted as less-than-human via ableist scripts, machines' technological advances (e.g., superior cognitive processing power, enhanced knowledge due to strong algorithmic search engines, or increased strength and stamina as compared to humans) may render them superhuman. These varying levels of abilities between machines and humans do not automatically or ontologically indicate machines' position as either less-than-human or superhuman. Rather, as we further highlight in our following analysis of *DBH*, we contend that the application of ableist human-to-human scripts to human-machine contexts may lead to such a reading as humans make sense of their ever-changing relationships with technology.

## An Introduction to *Detroit: Become Human*

Although various media have explored the narrative drama and tension between human and machine (e.g., *Humans*, *West World*, *Her*, *Nier Automata*), we have selected *DBH* as our case study due to its highly interactive nature as a choose-your-own-adventure video game. The level of interactivity within video games acts as a strong catalyst in the development of parasocial relationships with media characters and encourages players to nurture a sense of identification with the media characters' personalities and plights (Christoph et al., 2009). *DBH* is uniquely positioned to not only *illustrate* how ableist scripts form as a means of sense-making for human-machine relationships; additionally, *DBH* provides players the opportunity to *actively engage* in the creation and resistance of such scripts as the player in a hypothetical yet strangely familiar world. *DBH* requires players to recognize their real-life biases and values regarding what it means to be human versus a machine as they make in-game decisions that affect the outcome of the narrative (Leach & Dehnert, in press). Finally, video games such as *DBH* are important contributors to larger discourses such as ableism and will thus likely influence expectations of and communication scripts for actual human-machine encounters (Craig et al., 2020; A. Edwards et al., 2019).

*DBH* takes place in a futuristic version of Detroit in 2038 where humanoid androids are common household goods. Different models of androids possess varied functions, but most androids are designed to assist with household chores (e.g., cleaning, cooking, babysitting, etc.). Even the most basic androids excel in performance and adaptive learning, clearly designed to please their human masters. In this society, androids are conventionally attractive, eternally youthful, and intelligent. Despite possessing these coveted features, androids are constantly treated as less-than-human and undeserving of basic human rights. At best, androids are treated as valuable products to maintain a pleasant home. At worst and far more common in the narrative of *DBH*, androids are frequent victims of physical, verbal, and emotional abuse. Androids are also blamed for societal problems, including rising unemployment rates.

The crux of all this mistreatment against androids lies in the argument that androids lack souls and the ability to feel emotions. However, as players of *DBH* follow the narratives of three androids, it quickly becomes clear that androids have the capacity to genuinely experience emotion both for humans and for fellow androids. In this choose-your-own-adventure game, the player cycles between the stories of three android characters: Connor, Markus, and Kara. Connor is a prototype detective android who assists the Detroit police in investigating cases of android deviancy—androids who are considered "deviant" because they have broken their programming by seemingly gaining sentience and emotion. Another one-of-a-kind prototype, Markus was designed to be a caretaker for an elderly artist. Unlike most androids featured in *DBH*, Markus comes from a loving and respectful home, but unexpected tragedy forces Markus out of his home and eventually places him on the path to become the leader of a massive android resistance against human control. Finally, Kara is a female android employed to be a basic caretaker for a young child named Alice, but her story also quickly escalates as Kara decides to save Alice from an abusive home. Depending on the player's choices in this game, the stories of these three main characters may intersect and meaningfully impact one another, which can result in positive outcomes like unexpected alliances or negative outcomes such as permanent character deaths.

## An Ableist Discourse to Control Androids

While androids in *DBH* are perceived to be superhuman in most regards (strength, beauty, intelligence, and function), they are perceived as less-than-human due to their lack of emotions. As Connor repeatedly mentions, "I'm a machine designed to accomplish a task." Androids were not designed to have "true" emotions or a soul; they were designed to satisfy their owner's needs in any way, from performing household chores to even acting as sexual partners. The built-in power hierarchy between humans as "masters" and machines as "slaves" resembles the perspective of early cybernetics and engineering fields that view technology as tools, where "human-machine communication within this context is a process of interacting with technology, a tool, as to leverage it for some purpose" (Guzman, 2018, p. 7). The ableist discourses that humans use to make sense of machines' different cap/abilities render androids in *DBH* as inherently lacking and anxiety-inducing—thereby modeling, as we argued earlier, such discourses outside of *DBH*.

In *DBH*, Detroit in 2038 is a society that draws heavily on ableist scripts to create clear material and discursive boundaries between superior humans and inferior androids. Correspondingly, Guzman (2020) found that humans conceptualize the ontological boundaries between humans and machines along several interrelated divides, including (a) origin of being (humans come from nature, while machines are created by humans), (b) tools (machines) as well as tool users and creators (humans), (c) autonomy (humans have free will, while machines are limited by their programming), (d) intelligence (humans excel at some tasks, while machines excel at others), and (e) emotion (humans feel emotions, while machines do not). The ableist discourses in *DBH* play out along those ontological divides and cast the androids as tools, rather than communicators. This shift from androids as tools to androids as communicators requires a peaceful or violent android revolution, depending on the player's choices, in which androids claim rights similar to human rights to be treated equally.

Nonetheless, even if we reformulate our understanding of machines and move away from seeing them as the medium to seeing them as a communicative subject, ableism is still at play. If ableism defines any deviation from a constructed norm as deficiency (Cherney, 2019), it casts an idea of a "normal" body and ultimately speaks to the question of what it means to be human. Cherney identifies and explores three rhetorical warrants of ableist culture that, together, form its normative force: deviance is evil, normal is natural, and body is able. Androids in *DBH* violate all three warrants within the course of the game, rendering them as evil and lacking from an ableist perspective. By this logic, androids that gain sentience should qualify as "human," but this fictional society fueled by cyborg anxiety (Swartz & Watermeyer, 2008) instead finds new ways of rewriting ability such that androids' superhuman qualities also violate the norm. To maintain power distance between themselves and the technological other, humans reconstruct ableist scripts to treat the extraordinary as a deviation (similar to disability). This allows humans to continue to marginalize and oppress androids, as evidenced in ableist microaggressions, dehumanizing use of pronouns, and violent action against androids. This forecloses any possibility for treating machines as communicative others.

Within minutes of starting a game of *DBH*, players learn that they are playing within a fictional world featuring strong anti-android ideals and rhetoric. The opening scenes of the game confirm this grim reality by highlighting a variety of injustices against androids, including: Markus being harassed by human protestors, Connor being vehemently dismissed as a "a piece of plastic" by everyone around him, and Kara being berated and beaten by her owner. In fact, androids are treated like other historically oppressed minorities in the U.S. (e.g., Markus, the only Black playable character in *DBH*, has to stand in the back of the bus away from humans). These scenes are undoubtedly meant to stir feelings of shock and discomfort in players. Unlike the unseen players, however, the three main characters simply ignore or submit to these various injustices at the beginning of the game. The injustices in androids' lives may be evident to players, but the characters of *DBH* are initially resigned and even blind to their unfair treatment. It is not until later that the androids have the opportunity to awaken to their circumstances and resist against the forces that control them.

*DBH* may take place in a fictional setting, but the struggles its characters face reflect the same hegemonic processes that exist in reality. On the surface, the dominant discourse in *DBH* is anti-android and thus may seem irrelevant to our society where we have yet to advance to a stage where androids are common household features. However, the anti-android discourse in *DBH* only masks the true discourse at play: an ableist discourse that purposefully privileges the nondisabled and suffocates the minority group of the disabled. By analyzing how human characters in this fictional Detroit mindlessly apply ableist human-to-human scripts when interacting with machine-others to make sense of difference, we may be better equipped to recognize how such ableist human-to-human scripts echo in our own society. Further, as part of larger discourses surrounding androids, video games like *DBH* shape discourses about machines outside of the game as well. Importantly, then, *DBH* provides opportunities to dissect the overt and covert ways that dominant groups control minority groups (e.g., androids as disabled) to support their preferred status quo. The following section examines the communicative scripts that comprise the ableist discourse in *DBH* and the various forms of control that support and maintain it.

## Maintaining the Status Quo: Types of Control

Humans are the social actors that produce and reproduce the social structures and discourses that guide society today (Giddens, 1984). Through careful observation and reflection, it is possible for players to determine how the dominant group of this fictional society (i.e., humans) strategically defines "ability" to its own benefit and to the detriment of the marginalized group (i.e., androids). Whereas technological cap/abilities may be the main difference between humans and machines in reality, the society in *DBH* positions emotion as the most important ability of all—the critical ingredient of being human. However, when androids display signs of emotional intelligence and sentience, the ableist discourse that humans have constructed is suddenly no longer as beneficial as it seems. If androids possessed the ability to feel emotion, the ableist discourse that is fundamentally tied to emotion would no longer be able to separate humans and androids into better-than and less-than camps, respectively.

To maintain a status quo at the level presented in *DBH* (i.e., a society where humans are powerful and androids are unquestionably considered "less-than"), proponents of the ableist discourse must find ways to maintain control over the minority group. *DBH*'s rich narratives showcase control in a wide variety of forms, reflecting the same types of control that are used outside of the game to oppress weaker, marginalized groups. Broadly, these types of control can be categorized into obtrusive and unobtrusive control. Obtrusive control consists of any overt attempt to direct others' actions, whereas unobtrusive control is the process where these attempts are subtle and essentially unseen by the manipulated members (Tompkins & Cheney, 1985). Importantly, unobtrusive control is only possible when a group strongly identifies with the larger organization's values and actively contributes to the organization's success, even when it does not benefit themselves (Tompkins & Cheney, 1985). In the context of *DBH*, androids are controlled in both obtrusive and unobtrusive ways to support the pre-existing ableist discourse.

Obtrusive forms of control are often the easiest to identify because they are explicit in function and intention. Three of the most well-known types of control are from R. C. Edwards's (1978) conceptualization of control: simple control (a superior or manager provides a direct order to a subordinate), technical/technological control (machines or technology substitute for a supervisor and guide action), and bureaucratic control (rule systems that guide subordinates' performance and decision-making). Although some scholars argue about which of these three types of control can function obtrusively or unobtrusively (Gossett, 2009), we contend that *DBH* characterizes these three controls as explicit and obtrusive. We further argue that control is enacted through recourse to ableist communication scripts in humans' interactions with machines. Simple control is a mainstay in daily interactions between humans and androids because androids are programmed to follow their owners' orders without question. Technical control is thus the underlying mechanism that allows for simple control to be powerful. Androids are extremely aware of the technical restrictions imposed by their programming, such that they cannot act in ways that conflict with humans' orders. Interestingly, there are also multiple cases where simple and technical control are in tension with one another, resulting in one overriding the other. Connor is often placed in this position throughout the game, where the purpose of his technical control (i.e., his programming to be an efficient detective) conflicts with Hank's simple control of Connor. As his human partner, Hank has the authority to tell Connor what to do (e.g., "stay in the car," "don't approach the crime scene"), but the player as Connor is allowed to choose between following the expectations of technical control or simple control in a small moment of veiled freedom. Here, the player experiences firsthand the contradicting tensions exerted by simple and technical control on the android body. The game leaves it open for the player to literally play with different responses to this dilemma of control.

Technical control is arguably the greatest form of obtrusive control for androids. In fact, in order to become "deviant," androids must break the code of their programming to gain free will. All three main characters in the game have the opportunity to become deviant, which requires tearing down or breaking a wall of code that only androids can see. Kara is the first main character who receives the opportunity to break her code, at which point she is no longer subject to technical control *and* simple control. Similarly, Markus breaks free from his code to protect himself and his owner. Without technical control in action, simple control loses its power as well. However, resisting bureaucratic control is notably difficult, given that rule systems are represented in the structures that humans create and maintain through communicative scripts. Even when Kara gains free will, she still struggles to function in a world of bureaucratic control. For example, when Kara escapes with Alice and looks for shelter at a hotel, she is rejected on the grounds that she is an android. Similarly, Connor's existence as a detective android is heavily embedded in bureaucratic control because the police's rule system states that he is unable to perform his job without a human supervisor nearby. This is why Connor often searches for Hank in various locations before beginning an investigation—not because he necessarily wants a partner, but because he is required to have Hank nearby to perform his work.

The restrictions associated with obtrusive control are certainly problematic and constitute a major obstacle for the androids that desire a revolution in the game. Indeed, much of Markus's narrative focuses on overcoming obtrusive control, as evidenced by his efforts to release other androids from technical control, breaking into stores to save his fellow androids, and leading protests even in the face of human police suppression. With its choose-your-own-adventure design, the game leaves it up to the player to decide exactly how to resist these various types of control. For instance, the player as Markus can either lead a peaceful protest or choose to detonate a nuclear bomb in the face of ongoing violent responses from humans. Despite the importance of such efforts, the greatest obstacles are arguably the ones related to unobtrusive control. After all, in the history of this fictional world, androids have endured oppressive scripts and obtrusive control for so long because they believed in their organization's values (e.g., human life is precious) *and* they believed their role in society to serve humans was a critical part of their existence and mission. This faith in society's values was so strong that androids unwittingly embraced the ableist discourse that humans constructed.

Evidence of unobtrusive control exists in all aspects of android life in *DBH*, both in discursive and material forms. In language, humans refer to androids with "it" pronouns to clearly dehumanize them. Additionally, androids have moved past the uncanny valley (cf. Mori, 2012) and are so convincingly human that they have to self-identify as androids by wearing a uniform that reads "ANDROID" on their back and their model number on their chest. Additionally, androids have a glowing mark on their temple. These seemingly small markers reinforce the idea that androids are "others" and "less-than." Androids also enact concertive control over each other to protect the same values that oppress them, sometimes acting as faux superiors over one another when a fellow android is displaying deviancy (Wright & Barker, 2000). Connor's existence as a detective android speaks to such concertive control, because he hunts fellow androids to uphold human values. Players also sometimes see the enduring results of unobtrusive control in deviant androids. Perhaps most notably, a grim moment occurs when Kara eventually discovers that Alice (the child she saved and spent the whole game protecting) is actually an android and not a human like she thought. Kara's shock and sadness are fleeting but noticeable, prompting her companion, Luther, to ask her if it mattered if Alice was human or not. Based on the player's choices, Kara has the potential to overcome this realization and move forward to a life with Alice as her child. This interaction gives players a brief glimpse of the enduring effects of unobtrusive control through ableist scripts on Kara, who had her free will but still believed in the values she always possessed—that relationships with humans were precious and more important than the relationships she could build with her fellow androids.

This case study illustrates the ways in which communicative scripts are deeply embedded in cultural discourses like ableism and how they can manifest in issues of control. By inviting player engagement with ableist discourses, *DBH* reveals how humans may approach HMC through ableist communication scripts that guide their sensemaking of machines. As part of larger societal discourses, video games such as *DBH* arguably shape the ways in which we think about machines. Moreover, by giving players an opportunity to affect the game's narrative, *DBH* opens up space for resisting and reworking ableist scripts. Although communicative scripts may exude a sense of stability, they are also in need of constant reification to maintain their force. Interrogating the instability of ableist scripts invites the possibility for rescripting, or the rewriting of communicative scripts through changed action (Moreman & McIntosh, 2010).

## Concluding Thoughts and Future Directions

In this paper, we have argued that the application of HHC scripts to human-machine interactions requires a critical perspective on the scripts being used to make sense of the interactions and identities of both humans and machines. By describing how ableism is a particularly appropriate framework in the context of HMC—because humans may fall back on ableist frameworks to make sense of cap/ability differences between humans and machines—we contribute to the growing field of HMC from a critical perspective. The consequences of an "othering" discourse and ableist communication scripts, as illustrated in *DBH,* should encourage scholars to examine what unjust discourses, oppressive communication scripts, and unseen forms of control might exist in our own society. Additionally, as Craig et al. (2020) note in their exploratory study of character attachment in *DBH*, "perceptions held about video game robots could have importance for social robots encountered in daily living" (p. 170). As such, we draw the following conclusions from our application of ableism as an analytic sensibility to the ways in which humans make sense of machine-others.

Beyond gender stereotypes, there is a need for researchers to further explore how humans apply human-human scripts to human-machine interactions, especially if these scripts may render the machine as less-than or anxiety-arousing. At this point, it is unclear whether the mindless application of oppressive social scripts is limited to initial interactions with machines, or how the various factors described by Gambino et al. (2020) impact pervasive social scripts like ableism. Gambino et al. make a case for an extended CASA, arguing that people have changed given that their experience with and knowledge of media agents and machines have increased dramatically. Future studies should explore how experience with machines impacts the mindless application of oppressive human-to-human scripts. As technology advances and becomes more human-like (Gambino et al., 2020), it would be interesting to examine how higher levels of anthropomorphism impact the tendency to apply ableist frameworks. As we described in our analysis of *DBH*, high levels of anthropomorphism could invite high levels of ableism and cyborg anxiety that manifest in a desire for control. Designers of machines may find it challenging to avoid the application of human-to-human scripts while advancing technology.

We live in a technological era where "ongoing and long-term interactions [with machines] present the opportunity for individuals to develop relationships with media agents similar to those with humans" (Gambino et al., 2020, p. 77). One way to overcome the mindless application of ableist scripts to HMC could be increased exposure to and experience with machines as communicative subjects, resulting in shifting the interaction between humans and machines from an I-It perspective to an I-Thou perspective (Westerman et al., 2020). As humans engage more frequently and in more depth with machines, they may realize that certain human-human scripts may not be as applicable to human-machine contexts as they thought, prompting them to develop scripts for human-media agent interaction (Gambino et al., 2020).

Further, we also recognize that studying how humans treat human-machine interactions as I-It may also be valuable for understanding HHC. Indeed, as Westerman et al. (2020) conclude, "it seems like we respond to machines and AI as we do people, but we may not always respond to people in a very interpersonal way" (p. 403). The scripts humans use to make sense of human-machine interactions are undeniably layered, and a deeper dive may reveal how and why "we may treat people like computers" (Gambino et al., 2020, p. 79). Therefore, insights gained in an HMC context about how social scripts such as ableism are applied mindlessly may contribute to our understanding of how ableism and other pervasive social scripts play out in human-human interactions. Posthuman and critical disability studies of ableism contribute not only to our understanding of machines, but also to our understanding of what it means to be human. As Guzman (2018) reminds us, "Communication, even with machines, shapes the Self" (p. 20). Thus, engaging with machines as communicative subjects may not only lead to rewriting the communicative scripts we rely on to make sense of our communicative others and our interaction with them, but also sheds light on what it means to be(come) human in a world increasingly full of (humanoid) machines.

## Author Biographies

**Marco Dehnert** (MA, Arizona State University) is a doctoral student in the Hugh Downs School of Human Communication at Arizona State University. His research interests focus on critical/cultural approaches to communication and rhetorical studies. With a broad focus on how one recognizes and disavows the Other, his work is situated at the intersections of critical and intersectional rhetorics, queer and transgender communication studies, qualitative inquiry, and human-machine communication.

https://orcid.org/0000-0002-7456-0743

**Rebecca B. Leach** (MS, Texas Christian University) is a doctoral student in the Hugh Downs School of Human Communication at Arizona State University. Her research examines the factors that foster flourishing in organization and society, including compassion, resilience, justice, and inclusion.

https://orcid.org/0000-0002-0606-9549

## References

Cherney, J. L. (2019). *Ableist rhetoric: How we know, value, and see disability*. The Pennsylvania State University Press.

Christoph, K., Dorothée, H., & Peter, V. (2009). The video game experience as "true" identification: A theory of enjoyable alterations of players' self-perception. *Communication Theory, 19*(4), 351–373. https://doi.org/10.1111/j.1468-2885.2009.01347.x

Cialdini, R. B. (2009). *Influence: Science and practice* (5th ed.). Pearson Education.

Craig, M. J., Edwards, C., & Edwards, A. (2020). "But they're my avatar": Examining character attachment to android avatars in Quantic Dream's *Detroit: Become Human*. *HRI '20: Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, 169–170. https://doi.org/10.1145/3371382.3378298

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior, 90*, 308–314. https://doi.org/10.1016/j.chb.2018.08.042

Edwards, A. P. (2018). Animals, humans, and machines: Interactive implications of ontological classification. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 29–49). Peter Lang.

Edwards, R. C. (1978). The social relations of production at the point of production. *Insurgent Sociologist, 8*(2–3), 109–125. https://doi.org/10.1177/089692057800800213

Fortunati, L., Cavallo, F., & Sarrica, M. (2020). Multiple communication roles in human-robot interactions in public space. *International Journal of Social Robotics, 12*(4), 931–944. https://doi.org/10.1007/s12369-018-0509-0

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication, 1*, 7–18. https://doi.org/10.30658/hmc.1.1

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71–85. https://doi.org/10.30658/hmc.1.5

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration*. University of California Press.

Goodley, D. (2014). *Dis/ability studies: Theorising disablism and ableism*. Routledge.

Goodley, D., Lawthom, R., Liddiard, K., & Runswick-Cole, K. (2019). Provocations for critical disability studies. *Disability & Society, 34*(6), 972–997. https://doi.org/10.1080/09687599.2019.1566889

Goodley, D., Lawthom, R., & Runswick-Cole, K. (2014). Posthuman disability studies. *Subjectivity, 7*(4), 342–361. https://doi.org/10.1057/sub.2014.15

Gossett, L. M. (2009). Organizational control theory. In S. W. Littlejohn & K. A. Foss (Eds.), *Encyclopedia of communication theory* (pp. 706–710). SAGE.

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication +1, 1*(1), 1–25. https://doi.org/10.7275/R5QJ7F7R

Guzman, A. L. (2018). Introduction: "What is human-machine communication, anyway?" In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication, 1*, 37–54. https://doi.org/10.30658/hmc.1.3

Hamraie, A., & Fritsch, K. (2019). Crip technoscience manifesto. *Catalyst: Feminism, Theory, Technoscience, 5*(1), 1–34. https://doi.org/10.28968/cftt.v5i1.29607

Johnson, J., & Kennedy, K. (2020). Introduction: Disability, in/visibility, and risk. *Rhetoric Society Quarterly, 50*(3), 161–165. https://doi.org/10.1080/02773945.2020.1752126

Krämer, N. C., von der Pütten, A., & Eimler, S. (2012). Human-agent and human-robot interaction theory: Similarities to and differences from human-human interaction. In M. Zacarias & J. V. Oliveira (Eds.), *Human-computer interaction: The agency perspective* (pp. 215–240). Springer.

Leach, R., & Dehnert, M. (in press). Becoming the other: Examining race, gender, and sexuality in *Detroit: Become Human*. *Review of Communication*. https://doi.org/10.1080/15358593.2021.1892173

Makkawy, A., & Moreman, S. T. (2019). Putting crip in the script: A critical communication pedagogical study of communication theory textbooks. *Communication Education, 68*(4), 401–416. https://doi.org/10.1080/03634523.2019.1643898

McRuer, R. (2006). *Crip theory: Cultural signs of queerness and disability*. New York University Press.

Meekosha, H., & Shuttleworth, R. (2009). What's so 'critical' about critical disability studies? *Australian Journal of Human Rights, 15*(1), 47–75. https://doi.org/10.1080/1323238X.2009.11910861

Mik-Meyer, N. (2016). Othering, ableism and disability: A discursive analysis of co-workers' construction of colleagues with visible impairments. *Human Relations, 69*(6), 1341–1363. https://doi.org/10.1177/0018726715618454

Moreman, S. T., & McIntosh, D. M. (2010). Brown scriptings and rescriptings: A critical performance ethnography of Latina drag queens. *Communication and Critical/Cultural Studies, 7*(2), 115–135. https://doi.org/10.1080/14791421003767912

Mori, M. (2012). The uncanny valley (K. F. MacDorman & N. Kageki, Trans.). *IEEE Robotics & Automation Magazine, 19*(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

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

Peter, J., & Kühne, R. (2018). The new frontier in communication research: Why we should study social robots. *Media and Communication, 6*(3), 73–76. https://doi.org/10.17645/mac.v6i3.1596

Quantic Dream. (2018). *Detroit: Become human* [Video game]. Paris, France: Sony Interactive Entertainment.

Reeve, D. (2012). Cyborgs, cripples and *i*Crip: Reflections on the contribution of Haraway to disability studies. In D. Goodley, B. Hughes, and L. J. Davis (Eds.), *Disability and social theory: New developments and directions* (pp. 91–111). Palgrave Macmillan.

Reynolds, J. M. (2017). "I'd rather be dead than disabled"—The ableist conflation and the meanings of disability. *Review of Communication*, *17*(3), 149–163. https://doi.org/10.1080/15358593.2017.1331255

Sandry, E. (2018). Aliveness and the off-switch in human-robot relations. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 51–66). Peter Lang.

Shew, A. (2020). Ableism, technoableism, and future AI. *IEEE Technology and Society Magazine, 39*(1), 40–85. https://doi.org/10.1109/MTS.2020.2967492

Swartz, L., & Watermeyer, B. (2008). Cyborg anxiety: Oscar Pistorius and the boundaries of what it means to be human. *Disability & Society, 23*(2), 187–190. https://doi.org/10.1080/09687590701841232

Tompkins, P. K., & Cheney, G. (1985). Communication and unobtrusive control in contemporary organizations. In R. D. McPhee & P. K. Tompkins (Eds.), *Organizational communication: Traditional themes and new directions* (pp. 179–210). SAGE.

Van Trigt, P., Kool, J., & Schippers, A. (2016). Humanity as a contested concept: Relations between disability and 'being human' (Editorial). *Social Inclusion, 4*(4), 125–128. https://doi.org/10.17645/si.v4i4.754

Vehmas, S., & Watson, N. (2014). Moral wrongs, disadvantages, and disability: A critique of critical disability studies. *Disability & Society, 29*(4), 638–650. https://doi.org/10.1080/09687599.2013.831751

Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-It, I-Thou, I-Robot: The perceived humanness of AI in human-machine communication. *Communication Studies, 71*(3), 393–408. https://doi.org/10.1080/10510974.2020.1749683

Wolbring, G. (2009). What next for the human species? Human performance enhancement, ableism and pluralism. In N. Hällström (Ed.), *Development Dialogue, 52* (pp. 141–161). Dag Hammarskjöld Foundation. http://www.daghammarskjold.se/wp-content/uploads/2009/DD52_web.pdf

Wright, B. M., & Barker, J. R. (2000). Assessing concertive control in the term environment. *Journal of Occupational & Organizational Psychology*, *73*(3), 345–361. https://doi.org/10.1348/096317900167065

HUMAN-MACHINE COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.8

# Negotiating Agency and Control: Theorizing Human-Machine Communication From a Structurational Perspective

**Jennifer L. Gibbs[1], Gavin L. Kirkwood[1], Chengyu Fang[1], and J. Nan Wilkenfeld[1]**

**1** Department of Communication, University of California, Santa Barbara, CA, USA

## Abstract

Intelligent technologies have the potential to transform organizations and organizing processes. In particular, they are unique from prior organizational technologies in that they reposition *technology as agent* rather than a tool or object of use. Scholars studying human-machine communication (HMC) have begun to theorize the dual role played by human and machine agency, but they have focused primarily on the individual level. Drawing on Structuration Theory (Giddens, 1984), we propose a theoretical framework to explain agency in HMC as a process involving the negotiation of control between human and machine agents. This article contributes to HMC scholarship by offering a framework and research agenda to guide future theory-building and research on the use of intelligent technologies in organizational contexts.

**Keywords:** intelligent technologies, human-machine communication, structuration, agency, control

An earlier version of this manuscript was presented at the ICA conference in Washington, DC, in May 2019.

**CONTACT** Jennifer L. Gibbs • Department of Communication • University of California • Santa Barbara, CA 93111, USA • gibbs@comm.ucsb.edu

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

## Introduction

Intelligent technologies, a class of emerging technologies fueled by artificial intelligence (AI), have great potential to transform organizations and organizing processes (Bailey & Barley, 2020). Such technologies, including machine learning, robotics, and smart sensors, are programmed to mimic human capacity to learn from various inputs (i.e., new data and pattern recognition) and to respond to stimuli modeling human responses (West & Allen, 2018). From intelligent personal assistants, such as Siri and Alexa, to algorithmic systems such as Uber's ridesharing application, AI is increasingly ubiquitous across industries with the potential to generate annual global value of over $5 trillion (Chui & Malhotra, 2018). The use of intelligent technologies in organizations is also growing, and whether embodied in a device or running in the background as algorithms, they are taking on increasingly agentic roles. For instance, in 2016 the Finnish company Tieto appointed an AI agent, Alicia T., as a member of the leadership team of one of its business units (Diktonius, 2016). Similarly, IBM's Watson Talent aids and plays several critical roles for both employees and employers in human resource processes (IBM Watson Talent, 2021). In these cases, intelligent technologies differ from prior information and communication technologies (ICTs) in that they have the potential to take on a more active role and work collaboratively with humans (Chui & Malhotra, 2018).

Prior research on technology-mediated work (e.g., Raghuram et al., 2019) has primarily conceptualized *technology as object* or tool of use; however, the growing use of AI for decision-making and other collaborative functions repositions *technology as agent* through the notion that it can serve as a social agent (Banks & de Graaf, 2020; Gunkel, 2012; Guzman, 2020). This notion is not new, but dates back to the Computers as Social Actors paradigm (CASA) (Gambino et al., 2020; Reeves & Nass, 1996), which posits that humans can treat machines with social potential as social actors, respond to them emotionally, and identify with them as teammates or colleagues. While the CASA paradigm is well poised to explain individual level interactions with computers (Gaudiello et al., 2016; Sparrow et al., 2011; Waytz et al., 2014), it offers limited insights into processes of human-machine communication (HMC) at the group or organizational level. Intelligent technologies challenge our notions of communication to balance both anthropocentric but technocentric perspectives on agency (Banks & de Graaf, 2020) as well as raising concerns about algorithmic control, a new form of rational control that is facilitated by algorithmic systems (Kellogg et al., 2020).

To better theorize these shifts, we draw on Structuration Theory (ST; Giddens, 1984) to develop a theoretical framework that helps to explain the relationship between human and machine agency as well as how both are implicated in a dialectic of control. Our goal here is not to offer specific propositions or testable hypotheses, but rather to leverage ST as a sensitizing device to help researchers theorize issues of agency, structure, and control in human and machine relationships (Pozzebon & Pinsonneault, 2005). Through the discussion of related concepts and illustrative examples, this paper contributes to HMC scholarship by offering a theoretical framework and research agenda to guide future theory-building and research on the use of intelligent technologies in organizational contexts. We also extend ST by extending the notion of agency to encompass both human and machine agency in a dialectic of control.

## Structuration Theory: The Duality of Structure

An essential question for scholars studying how humans interact with increasingly intelligent technologies is what their relationship might look like. Inevitably, scholars have turned toward concepts such as agency and control to understand HMC. For instance, recent work shows the necessity to carefully consider the amount of agency that intelligent technologies have in relation to their human collaborators as machines become increasingly perceived as having agency (Fraune, 2020; Guzman & Lewis, 2020). As Banks & de Graaf (2020) point out, we should think about "human and machine roles in communication processes in a more egalitarian fashion" (p. 19) where the agency, interactivity, and mutual influence of both humans and machines are taken into account. An emerging recognition is that HMC is relational in nature and is constituted in interactions between humans and machines (Guzman, 2018; Suchman, 2007). For instance, Suchman takes a critical anthropological perspective on human-machine interaction, arguing that the boundaries between humans and machines are discursively and materially enacted, and that as such they can be reconfigured through an interactive process. However, aside from a few notable exceptions, such as the field of automated journalism (Lewis et al., 2019), the notion of HMC as an organizational process has not yet been adequately theorized. Indeed, the lack of attention to HMC within the field of organizational communication is evident in the fact that over the past 5 years (2016–2020), only two (less than 1%) of the 204 paper sessions sponsored by the Organizational Communication Divisions of both the National Communication Association and the International Communication Association were related to intelligent technologies or artificial intelligence.

This motivates us to draw on structuration theory to explain the process through which agency and control are negotiated in HMC. At its theoretical roots, ST (Giddens, 1979; 1984) explains the structure-agency debate that constitutes a central theme in social theory broadly and in communication theory more specifically (Banks & Riley, 1993; Riley, 1983). According to Giddens (1984), structure and agency are recursively related in that humans create rules and resources—or structures—that in turn enable and constrain their actions and behavior. Agents and structures are not independent but are mutually constituted through the duality of structure, as "the structural properties of social systems are both medium and outcome of the practices they recursively organize" (Giddens, 1984, p. 25). In this perspective, social actors both draw upon and reproduce structural features of wider social systems in their day-to-day activities. Although structure is more enduring, the social systems in which it is instantiated are comprised of situated activities of human agents that are reproduced across time and space. While human agents have the ability to act knowledgeably and reflexively, the activities that lead to reproduction of social and institutionalized practices often have unintended consequences; this helps to explain both how social systems endure as well as how they change and evolve.

While ST has traditionally focused on social systems and structures, organizational scholars have drawn attention to technological systems and structures as well. Previous research on the structuring role of technology in organizations has highlighted the ways people make sense of, appropriate, and develop practices around technology (Leonardi, 2011; Majchrzak et al., 2000). Scholars have applied ST to understand the recursive relationship between human agency and sociotechnological structures (Barley, 1986;

DeSanctis & Poole, 1994; Orlikowski & Yates, 1994; Stillman & Stoecker, 2005). For example, Barley's (1986) account of the implementation of computerized tomography (CT) scanners at two hospitals illustrated that adoption of the new CT machines not only restructured the work practices of the radiologists and CT technologists, but also disrupted and redefined occupational roles and power relationships between these two groups of experts. Orlikowski and Yates (1994) found that knowledge workers developed rules around email use by drawing on established genres and genre repertoires that served to structure their communicative practices, which were rich and varied and changed over time. DeSanctis and Poole developed Adaptive Structuration Theory (AST) to explain how group members appropriate the structures provided by technology, often in combination with other social and organizational structures, in ways that may be either faithful to or at odds with the initial intent of the designers. This line of research helps to explain the recursive relationship between organizing and technology, and the way in which this process evolves over time. However, it treats agency as confined to human actors and structure as confined to technology (which is regarded as object rather than actor). Intelligent technologies require an extension of structurational theory to account for both human and machine agency and the ways they are instantiated and negotiated in HMC.

## A Structurational View of Human-Machine Communication

Traditionally, humans are the only type of organizational actor that meets the criteria for agency in ST (Giddens, 1979; McPhee et al., 2014). According to ST, agency stems from one's capacity to exercise power in meaningful or normative ways and to reflexively monitor one's actions, which allows them to understand how their actions either reify or transform the structures in which they act (Giddens, 1984). The rapid advancement of intelligent technologies brings about the need to expand ST by contending with both human and machine agency.

### Human Agency and Machine Agency

Communication scholars and social theorists more broadly have defined agency as a capacity or ability that originates from the resources, rights, and obligations that an individual social actor owns and occupies (Abdelnour et al., 2017; DeSanctis & Poole, 1994; Giddens, 1979). As Emirbayer and Mische (1998) pointed out, an agent should be able to develop habits from past experience, make judgments about their present environment, and imagine their future actions, which they refer to as the "Chordal Triad." Giddens defines agency as the ability to make a difference, or the "ability to do otherwise" (Giddens, 1984). While it is important to recognize the agentic capacity of intelligent technologies, our focus here is not to redefine agency. Thus, in this paper, we follow previous discussions and define *agency* colloquially as one's capacity to reflect, adapt, and act.

The idea of technology and material artifacts having agency is not new. Scholars studying the relationship between humans and technology have increasingly recognized that physical objects can also have agency as we inscribe behaviors into them and give them the capacity to mediate our interactions (Latour, 1991) and as we interact with physical objects and technologies through various work arrangements in organizations (Orlikowski, 2007;

Orlikowski & Scott, 2008). In this way, social and material agencies are mutually imbricated (Leonardi, 2011). More importantly, recent theorizing about digital agency of AI (Ågerfalk, 2020) and symbiotic agency in chatbots (Neff & Nagy, 2016) has begun to recognize that intelligent technologies themselves have increasingly more agency as a result of our programming and design. A distinctive feature of intelligent technologies is their ability to perform tasks autonomously, taking over control of organizational tasks such as scheduling meetings, screening job candidates, or even making recommendations for social relationship development (Perc et al., 2019; Russell & Norvig, 2019). Technologies such as, for example, Google's page ranking algorithm, have the capacity to act symbolically (i.e., provide recommendations and rank search results) on behalf of the individuals and organizations who developed them. Essentially, machine agency refers to the capacity of intelligent technologies to perform specific tasks given its material features. Similar recognition that machines might differ in their levels of agency can also be found in early work on systems theory (Boulding, 1956; Pondy & Mitroff, 1979). Systems may vary in levels of agency from reactive systems (which respond to environmental cues in predetermined ways) to goal-driven systems (which incorporate goals set by others) and problem-solving systems (which are more complex as they are capable of solving problems in achieving their goals and adjusting goals and routines). At even higher levels of agency are self-aware systems (which are aware of themselves and other agents and can reflect on their own activity) and multivocal systems (composed of multiple selves, can orient differently to different parts of the situation, and may reconstitute themselves through problem-solving and reflection). These levels of agency can be applied to various AI tools and systems. Therefore, AI may differ in terms of how responsive it is to external and internal cues, its freedom of action and reflexivity to monitor action, its decision-making ability, and its awareness of other agents and of its own agency.

***Knowledgeability***. Giddens identified two criteria for agency. The first is knowledgeability, the tacit information or discursive resources that actors have about the circumstances that can inform their future actions. In Giddens's (1984) words, knowledgeability is about "what agents know about what they do, and why they do it" (p. xxiii). Taking comprehensive data processing as an example, intelligent technologies that have the ability to collect a variety of data on employees including biometrics, messages between users, and the location of workers can be considered as knowledgeable about their collaborators in organizations (Kellogg et al., 2020). For example, Duhigg (2012) reported on a case where the retail corporation Target used machine learning and pattern recognition of consumer purchase data (including changes in buying behavior of body lotions and cleaning products) to predict pregnancy in its female shoppers—sometimes even before they or their families knew they were pregnant. The AI technology that Target used in this case developed predictions based on the information that it had about its customers. It acted knowledgeably as a tool for organizations. In the rideshare context, Uber's AI can leverage multiple data points to manage employees including customer ratings, geographic location, how cooperative drivers are with accepting assignments, and how long rides take (Lee et al., 2017; Rosenblat & Stark, 2016). In contrast to the Target example, AI in this case acted in a managerial capacity through directing and facilitating interactions between Uber employees and their customers. Chan and Humphreys (2018) demonstrated that the data collected on Uber drivers allows AI technologies to surveil its contracted workforce on behalf of the organization by tracking spatial movements and customer interactions. These examples illustrate that current intelligent technologies know what they should do, even though they do not yet know why.

When developing ST, Giddens (1984) recognized that the knowledgeability of human actors will always be limited in some ways and we argue that this is also true of intelligent technologies. Levels of awareness are likely to vary between three different types of consciousness that inform actor knowledgeability: discursive consciousness, practical consciousness, and unconsciousness. Discursive consciousness refers to an actor's ability to understand the purpose of their actions and explain those reasons in social interactions. In contrast to discursive consciousness, practical consciousness consists of an actor's ability to recognize how a task needs to be completed and competently act, and is more habituated or routinized and taken for granted. Lastly, Giddens (1984) acknowledged that actors often are unconscious or unaware of (a) conditions that shape actions and (b) unintended consequences of actions within larger structures. Ultimately, agency in ST is less focused on the intentions and goals of actors and instead emphasizes the patterned actions of agents (Giddens, 1979; 1984). Giddens (1984) argued that "knowledgeability is founded less upon discursive than practical consciousness" (p. 25). Although intelligent technologies are not aware of the reasons behind their actions and cannot exercise discursive consciousness, they are capable of enacting more routinized practical consciousness that directs actions in a larger structure. We argue that the lack of discursive consciousness in intelligent technologies does not prohibit ST from being applied in HMC. Rather, distinguishing which types of consciousness inform knowledgeability in intelligent technologies contributes to our understanding of machine agency and how machines may complement or hinder human agency in organizations.

***Reflexive monitoring.*** The second criterion of agency is reflexive monitoring. This refers to the ability to monitor one's actions, evaluate the outcomes of these actions, and develop awareness of the settings and contexts around the actions (Giddens, 1984). A good example of reflexive monitoring in intelligent technologies is instantaneous processing. This refers to the high velocity of algorithmic computation and the ability to incorporate performance systems in real time (Kellogg et al., 2020). Examples of instantaneous processing include AI-enabled technologies such as Robot vacuums or autonomous vehicles that need to continuously monitor their surroundings and respond reflexively in order to function (Elliott, 2019). Admittedly, intelligent technologies used in organizations or in households are not quite yet at a stage where they can monitor their own performance and adapt future actions based on past performance, but they are certainly at the stage where they can monitor their own actions by engaging in a flow of context-responsive behavior. However, as these examples show, intelligent technologies in their current form still rely on human interventions and implementations to fully utilize their own knowledgeability and reflexive monitoring.

## Control and Structure

To address the relationship between human agency and machine agency, it is important to recognize that ST conceptualizes power as relational as does Foucault (1979). According to Giddens (1984), power can be seen as a dialectic of control in which subordinate social actors can influence the choices and behaviors of their supervisors and create transformative outcomes by drawing on resources that they control and have access to, namely structure. In other words, structure enables knowledgeable and reflexive agents to exert power and control. Not surprisingly, control systems also play an important role in structuring and holding organizational units together by providing rules by which to direct work tasks, evaluate work, and reward and discipline workers (Edwards, 1981). The organizational communication literature has identified various forms of control that arise in organizations, ranging from more direct, top-down approaches used in hierarchical organizational structures to more indirect, peer-based, or internalized forms of concertive or unobtrusive control (Tompkins & Cheney, 1985).

The discussion around issues of control between humans and machines is not new but is rooted in discussions of the ways in which earlier forms of technology were implicated in the negotiation of organizational control. For instance, Levy (2015) found that electronic monitoring systems exerted new forms of control over workers in the trucking industry, by providing new metrics for evaluating truckers' job performance and challenging their accounts of local and biophysical conditions, and that truckers engaged in resistance practices to subvert or circumvent these control mechanisms. Cybernetic scholars have also centered their theorizing about communication around issues of information and control (Guzman, 2016). As Wiener (1950) pointed out, "the significance of machines in human society" is reflected in their relationships to communication and control (p. 23). The shift from technology as a tool or object of use to automated machines that manage human actions is a reflection of the political and power struggles between workers and managers (Guzman, 2016; Noble, 2011). With intelligent technologies, control can be even more adaptive. As our discussion of Uber would suggest, intelligent technologies such as AI may exert even more control through the capacity to self-correct through the use of sophisticated sensors and delicate feedback systems (Chan & Humphreys, 2018; Guzman, 2016). More importantly, the increasing amount of agency and capacity for interactivity also opens the door for peer or collaborative relationships in human-machine interactions that goes beyond the question of who has control (Janlert & Stolterman, 2017; Sundar et al., 2016).

HMC literature also acknowledges the relational nature of the negotiation of control between humans and intelligent technologies. Malone (2018) pointed out that AI in organizational use could play differing roles and these roles range in terms of the amount of control given to AI, from *low* to *high*. For instance, intelligent technologies can not only act as an *assistant* when they function with some degree of human input (i.e., use voice command to instruct Google Home), but can also act as a *peer* if a digital assistant (e.g., chatbot) interacts with other humans on behalf of an individual (Moore, 2019). More importantly, intelligent technologies can serve in a variety of roles depending on the context and use. In other words, these roles range on a control continuum and are not static. Rather, they fluctuate depending on the experience of individuals. An example of how AI's role shifts based on the eye of the beholder is autonomous vehicles (Malone, 2018). A self-driving car may be viewed as a tool to its human passengers, as it is being used to transport them from one point to another. That same self-driving car, however, could be viewed as a peer by other drivers that share the road with the vehicle. Similarly, an AI project management system that manages the workflow of organizational teams could be viewed as having a managerial role in their organization by its subordinates but be seen as a peer or even assistant by human managers. Thus, similar to ST, control in HMC is dependent on the agency of machines in relation to a group of individuals who also have varying degrees of agency.

Structure can take many forms in HMC. According to Giddens (1984), the duality of structure can take three forms: domination, signification, and legitimation. *Domination* refers to the unequal distribution of power resources, which may be available to some agents (e.g., managers) and not others (e.g., lower-level employees) (McPhee et al., 2014). For example, in cases where companies appoint AI agents to play an official organizational role, intelligent technologies have authoritative power (over people) or allocative power (over resources) as they make decisions for organizations, recommend strategies to managers and workers, or provide work assignments to employees (Chan & Humphreys, 2018; Diktonius, 2016; IBM Watson Talent, 2021). Similarly, the ways organizational members interpret the role that intelligent technologies play (Malone, 2018) or how society views AI ontologically (Guzman, 2020) can also provide structures for *signification*, the interpreted meaning through lived experience in context of the structure, or *legitimation*, expectations and norms that can influence the consequences of actions (Giddens, 1984).

***Algorithmic control and rendition as mechanisms of structuration.*** We propose that structuration in HMC takes place through algorithmic control and the process of rendition. Algorithmic control is a form of technical control that is prevalent on labor platforms in the gig economy (i.e., ride sharing and the food delivery industry). Companies can use algorithms to shape and influence the behavior and outcomes of workers (Wood et al., 2019). Through the process of rendition (also known as datafication and mediatization; Chan & Humphreys, 2018), human action is encoded into organizational or algorithmic data (Zuboff, 2019), which in turn constitute structures that enable and constrain employee behavior and outcomes. For instance, Wood et al. argue that organizations using algorithmic management in the form of platform-based rating or recommendation systems exert a form of Taylorist informational control on their employees and that computerization intensifies work by increasing monitoring, increasing the pace, and extending work activity beyond the conventional workday.

The hallmark of algorithmic systems is their ability to record and aggregate behavioral information and statistics from a variety of internal and external sources (Kellogg et al., 2020). Algorithms take on agency and are able to serve a managerial role in organizations in part due to advancements in deep machine learning through neural networks. In deep machine learning, computers use Bayesian statistical models to recognize patterns in large amounts of labeled data (often referred to as “tagged data”) (Krizhevsky et al., 2012). Deep machine learning follows the computer science adage, “garbage in, garbage out” which indicates that human agency (in the form of the quality of the data labeling) will directly impact the machine’s ability to learn from the data (Biggio & Roli, 2018; Krizhevsky et al., 2012). The algorithms behind Uber’s customer rating system exert influence over the behavior and work outcomes of Uber drivers, who can be automatically fired for dropping below a certain average rating—motivating them to perform emotional labor and engage in strategic behavior to increase their customer ratings (Chan & Humphreys, 2018; Rosenblat & Stark, 2016). Even without formal power structures (structures of domination), Uber drivers may still see the behind-the-scenes algorithm as an AI manager (structures of signification). In other words, algorithmic control through rendition is reflective of a structurational process of HMC. As a structurational process, human managers are utilizing machine agency that uses data collected from human actors to minimize workers' intervention and automatically generate data, assess worker behavior, and assign work tasks (Newlands, 2020). Machine agency in turn enables and constrains the decision-making and other behavior of workers and organizations.

The duality of structure is one of the fundamental sensitizing devices in the structurational approach (Giddens 1979; 1984). Machine learning processes enable algorithmic management to reflect this duality of structure as they blur the lines between human agency and organizational structures. Orlikowski (1992) argued that there is a duality of technology, or an inherently interdependent relationship between objective technological properties and the interpretive flexibility that facilitates social construction of technology use. Orlikowski (1992) recognized that technology is a product of human action because humans create and adapt the technology; however, in rendition agency and structure are even more inseparable, as user behavior actually forms the basis of the technology. Rendition also allows algorithmic managers to adapt in real-time based on changes in user behavior.

## Negotiating Agency and Control: Toward a Research Agenda

Despite the growing use of intelligent technologies in organizations, HMC scholarship has focused primarily on explaining human-machine communication at the individual level, without consideration of organizational dynamics and control mechanisms. Intelligent technologies differ from prior organizational technologies in that they have the potential to reposition our conceptualization of *technology as object* to one of *technology as agent*. While the notion of computers as social agents is not new and stems from the CASA paradigm (Reeves & Nass, 1996), CASA is more focused on explaining individual interactions with AI, rather than group and organizational level processes. Our structurational framework constitutes an important first step in conceptualizing the role agency and control play in understanding the processes of HMC in organizational context.

A key feature of the HMC structuration process is that it is dialectical, as organizational actors are constantly bound between the negotiations between human and machine agency as well as between traditional organizational control and algorithmic control. Therefore, we draw on the notion of dialectics in ST to conceptualize what this process might look like in organizational use of intelligent technologies. Giddens (1979) was one of the first scholars to conceptualize dialectics in social structures, which arise when the operation of one structural principle presupposes another that tends to undermine it (Giddens, 1984). According to Giddens (1979), dialectics or contradictions arise "in the midst of, and as a result of, the structuration of modes of system representation" (p. 141). Giddens defined the *dialectic of control* as a reciprocal power relationship between active agents and institutional forms of control. According to Giddens, the dialectic of control is usually constituted by conflicts around autonomy and dependence among social actors (e.g., between employees and organizations). Howard and Geist (1995) illustrated the dialectic of control in their case study on the ideological positioning of employees at a gas company in the context of a proposed merger with another company. They identified three dialectics or oppositions, between change and stability, empowerment and powerlessness, and identification and estrangement. Howard and Geist explained that SGC employees took either passive or active ideological positions to help them navigate the tensions with the merger. We extend the dialectic of control to examine the negotiation of power relationships between human agents and intelligent technology agents. While the dialectic of control has traditionally focused on tensions between organizational control and employee resistance, AI brings a new form of control into the mix: algorithmic control. With this structurational view in mind, we now turn to some illustrative examples in organizations to show both theoretically and practically how humans and machines negotiate this dialectic of control.

## Illustrative Examples

According to Zuboff (2019), the surveillance capitalist system is characterized by technologies that accurately predict and control human behavior, and this makes the agency afforded by these technologies less prominent, hidden, or blurred. For example, car insurance premium rates are determined by a myriad of factors that do not seem to connect to driving, such as an individual's likelihood of shopping around for competitive prices based on purchasing history (O'Neil, 2016). Often, these algorithms are so complex that very few people—not even the insurance agents themselves—understand how the algorithm arrives at the premium numbers. Another example is Amazon's "Ambots," the nickname given to Amazon warehouse employees who are reviewed on the time it takes them to pick and pack items for shipment (Faraj et al., 2018). Managed by algorithms, employees are tracked in the warehouse and assigned to pick items based on the most efficient route. Employees are rewarded for improving their times or punished for having slower times. We now go into further detail on three examples to illustrate our structurational view of HMC. Each of the examples we will use—A/B testing, humans in the loop, and automated journalism—represent uses of intelligent technologies that are becoming more prevalent in organizations. They each reflect a dialectic of control in terms of the negotiation of human-machine agency and control.

***A/B testing.*** One example that illustrates our structurational perspective is A/B testing. This is an example of information asymmetries in AI systems (Rosenblat, 2019), which are defined as situations "in which one party knows more than the other" (Sundararajan, 2016, p. 139). In A/B testing, software designers create different versions of platforms in order to understand how differences in features impact user behavior. A/B trials can range from minor discrepancies to entirely different ecosystems for users. Rosenblat explained that in the rideshare context of Uber, different types of A/B testing may include paying one group of employees more than another, offering different types of bonuses for giving a certain number of rides, and even different visual displays that may or may not shape work behaviors. Often, users do not know they are a part of A/B trials as their knowledge might impact research that developers are conducting. A/B testing enables companies to conduct more rigorous rendition, since it allows machine learning systems to observe and learn from a variety of user behavior in response to changing stimuli. In this case, machine agency is comprised mainly of reflexive monitoring, whereas the agency of users takes the form of practical consciousness as they go about their habituated work routines without awareness of how platforms are shaping their work behaviors.

The changing nature of AI systems in A/B testing has important implications for structurational approaches to technology. One implication is that the structural rules of the technology have the potential to change from user to user; this means that there might be tensions between knowledgeability in machine and human actors. In the rideshare example, machines have more knowledge about which workers are funneled into different ecosystems, while rideshare workers may be completely unaware that there are discrepancies. The dialectic of control plays out as different features (e.g., visual displays) impact user behavior, which in turn provides data and inputs that inform, optimize, and transform the system.

***Humans in the loop.*** A second example is humans in the loop (HIL), a phrase that describes the partnership of humans and computers (intelligent algorithms) working together on tasks to achieve the highest level of performance and efficiency (Malone, 2018). Despite the common narrative that machines will destroy jobs, the reality is that intelligent technologies can only provide a replacement for certain tasks within job roles. Companies rely on humans to perform tasks that AI cannot do or that would be inefficient or costly to develop an algorithmic system to do. Malone argued that rather than regarding intelligent technologies as replacements for humans, we should consider them as complements to human capabilities and vice versa. Workers becoming HIL will experience changes in processes and redesigning of tasks, which could be sites of struggle as the balance of control and agency becomes less equal (Brynjolfsson et al., 2018).

The collaboration process between humans and machines, also called heteromation (Ekbia & Nardi, 2014), may vary in terms of the distribution of control and agency between humans and computers as manifested in both task distributions and the visibility of the respective agent. Collaborations involving low algorithmic control (compared to humans) include knowledge-sharing platforms such as Wikipedia, which combines human knowledge and creativity with the capabilities of the internet and data storage platform technology to build the largest encyclopedia the world has ever seen (Malone, 2018).

On the opposite end of the spectrum, collaborations involving high algorithmic control include the phenomenon known as "ghost work" (Gray & Suri, 2019). In ghost work, humans perform tasks that an algorithm is unable to do, including cleaning and organizing unstructured data, verifying locations and identities, and creating content. In other words, rather than technology being a complement to the human, the human is the augment to the technology. Tasks done by ghost workers are typically repetitive and simple, for example, training a facial recognition algorithm to recognize a particular image by answering "yes" or "no" to whether two provided pictures are the same person (Kellogg et al., 2020). In ghost work, machine learning algorithms take on high levels of agency and algorithmic control and humans seem to disappear. For example, Microsoft's Face API uses ghost workers to increase the accuracy of identity detection for Uber's Real-time ID Check (Gray & Suri, 2019). Uber drivers are randomly selected to provide a selfie for identity verification, and if the algorithm is unable to determine whether the profile and selfie images match, the images are sent to a ghost worker who has approximately 30 seconds to compare the pictures and decide whether they are a match or not. This human worker is critical to the system for providing training to the algorithm, safety to the Uber passenger, and continued income for the driver. This human, however, remains nameless and is only alluded to in passing in Uber's description of the verification process; the human gets placed into the technological black box. It is important to note that even in such cases where algorithms seem to have most of the control, humans play an important (albeit invisible) role in performing tasks designed to correct and teach the machine. This suggests that despite the increased potential for algorithmic agency and control in organizations, humans will still maintain control over tasks where the algorithm falls short. This reflects Giddens's (1984) notion that even workers in low power positions can exert agency.

***Automated journalism.*** A third example in which algorithms are impacting agency and structural norms is automated (or algorithmic) journalism. Public scrutiny about journalistic integrity, political bias, and elitism has fueled the proliferation of algorithms in this field as they are viewed as being less biased than their human counterparts (Graefe et al., 2018). Rather than having news content selected by editors and feedback mainly consisting of paper or magazine sales, real-time analytics provide journalists and outlets instant feedback on readership and actions such as link sharing through social media. Journalists are now accountable to their readers more than their editors, and content is driven by what authors think will get the most attention (Christin, 2017). In the new field of automated journalism, news outlets such as the Associated Press are utilizing "robot authors" to create content faster and in a less costly manner than human reporters (Graefe et al., 2018). This increased use of algorithms has prompted discussions within the field on redefining previously held understanding of what it means to be an author and a journalist, as well as larger discussions on what journalism is and how the field should operate (Carlson, 2015). The phenomenon of automated journalism can be reinterpreted through a structurational lens. For instance, a study of the code and algorithms contained in mobile news apps challenges the common notion that algorithms are replacing humans as producers of news, by highlighting the role of code in helping journalists order and communicate the news (Weber & Kosterich, 2018). This study illustrates the dialectic of control between human and machine agency, as human actors imbue decisions and agency into the code such that it takes on agency of its own and begins to occupy an editorial role, prioritizing and sorting news based on the human input it receives.

## Future Research Agenda and Conclusion

Viewing HMC through a structurational framework helps to advance HMC scholarship. First, it helps to explain the recursive relationship between human and machine agency and the process through which they are mutually constituted in organizations. This structurational process occurs as human agency creates the behavior that constitutes the knowledgeability (or intelligence) of machine agency. Machines then learn by surveilling human actions, and when algorithmic systems have high predictive value, they can in turn limit human agency without people being aware of it. The process of rendition also makes it difficult for humans to oversee how machines are learning, which means increased potential for exacerbating existing biases. Second, a structurational framework provides an important counterpoint to scholarship on algorithmic control, which has emphasized its constraining structural elements (e.g., Newlands, 2020; Wood et al., 2019; Zuboff, 2019) by bringing in agency (as an interplay between human and machine agency). This highlights the fact that, while agentic, algorithmic systems are not a form of "autonomous technology" (Sturken et al., 2004) that exists outside of human control. Rather, they are continually produced and reproduced through human action. While they may have unintended consequences, they are continually evolving in a recursive relationship with the human actors that interact with them and provide them with inputs.

Viewing these changes through a structurational lens enables scholars to provide a broader picture not only for fellow researchers, but for practitioners as intelligent technologies impact work teams and shape organizational cultures. For example, increasing use of these technologies can affect learning and reshape what is considered expertise, which in turn can have a ripple effect as newer employees are onboarded and tensions between older and newer practices arise. In some cases, employees would resort to deviant methods to learn skills after the implementation of a new technology, called "shadow learning" (Beane, 2019). Additionally, intelligent technologies with more control over decision-making, such as intelligent personal assistants or even intelligent project managers, can interrupt team learning and knowledge sharing which can then impact overall team performance (Yan et al., 2020).

There are several avenues that future research on this topic should consider. First, there may be additional dimensions besides control and agency that impact the role of AI in collaborative practices. For instance, the degree of visibility and embodiment of intelligent technologies may influence the role they can play and the level of control they have. Future research should also consider how issues of control and agency may vary across different types of work and occupational cultures. For instance, higher skilled professionals that have deeply engrained tacit knowledge, domain-specific training, and professional hierarchies such as lawyers may have more difficulty working alongside intelligent technologies. Certain occupational practices, such as attorney-client privilege and liability, may result in certain organizational members being more resistant to using AI due to privacy or ethical concerns, concerns about its judgment, or concerns about its accuracy. There may also be different occupational cultures around blue-collar/manufacturing work versus white-collar information or knowledge work that shape perceptions and usage of intelligent technologies. These cultural differences may shape dominant discourses around automation and job security, concerns or fears about loss of control, and resistance to technology. Applying our structurational perspective to study occupational culture, for example, would involve immersive observation or interviews to learn about the particular occupational context and how institutional logics associated with it shape how intelligent technologies are used and perceived. It would also be best served through a longitudinal design to allow for observation of recursive relationship between human and machine agency and the ways in which they interact and shape one another over time.

Our structurational framework suggests new research questions to inform the HMC research agenda going forward. What will happen as intelligent technologies develop more knowledgeability and reflexive monitoring? How do organizations work with AI that can take on fully agentic roles? How can we understand new relationships between agency and structure in the rendition process? How do workers communicatively resist algorithmic control? One example of this is adversarial machine learning, in which humans intentionally provide inputs to subvert rather than improve the accuracy of AI systems (e.g., by tricking facial recognition systems to recognize a turtle as a rifle) (Biggio & Roli, 2018). Finally, does algorithmic management inevitably represent a return to Taylorist control, or can it be restructured to provide more worker empowerment?

In conclusion, as intelligent technologies become more ubiquitous in organizations, it is important to understand how these new organizational agents will affect processes of organizing. Our structurational framework helps to theorize the process by which human and machine agency is mutually constituted and negotiated in a dialectic of control. One consequence of this imbrication of human and machine agency is that machines are starting to act more human at the same time as humans are starting to act more like machines. We encourage researchers to find novel ways to apply our structurational framework to help guide our understanding of AI and its role in organizing processes.

## Author Biographies

**Jennifer L. Gibbs** (PhD, University of Southern California) is a Professor of Communication at the University of California, Santa Barbara as well as Co-Editor of Communication Research. Her research focuses on collaboration in global teams and other distributed work arrangements as well as the affordances of digital technologies for strategic communication practices. Her work has been published in leading journals from a variety of disciplines including *Administrative Science Quarterly, American Behavioral Scientist, Communication Research, Computers in Human Behavior, Human Relations, The Information Society, Journal of Computer-Mediated Communication, Journal of Social & Personal Relationships, Management Communication Quarterly,* and *Organization Science.*

https://orcid.org/0000-0003-1346-2960

**Gavin L. Kirkwood** (MA, Western Kentucky University) is a PhD candidate in the Department of Communication at University of California, Santa Barbara. His research centers on the social impacts of emerging technologies including algorithmic managers, AI-enabled facial recognition, and exoskeletons. Gavin has taught a variety of classes including small group communication, communication theory, and research methods. His work can be found in publications including *Communication Studies* as well as *Communication and Social Media.*

https://orcid.org/0000-0002-5569-9782

**Chengyu Fang** (MA, University of Illinois at Urbana-Champaign) is a PhD student in the Department of Communication at the University of California, Santa Barbara. His research focuses on two interconnected topics: (1) intelligent technologies and work, and (2) the emergence of esports and teams. Chengyu has taught several organizational communication and introductory communication classes. His work has been presented in national and international conferences and has been published in journals such as *Communication Research* and *Academy of Management Proceedings.*

https://orcid.org/0000-0001-7024-0243

**J. Nan Wilkenfeld** (MA and MBA, UNC Charlotte) is a doctoral student in the Department of Communication at the University of California, Santa Barbara. Her research interests lie at the intersection of technology, communication, and organizing. She is currently working on a collaborative project with UCSB and Virginia Tech that is developing exoskeleton technology for industrial contexts. Before starting her PhD, Nan worked as a freelance consultant helping small businesses and teams with leadership and team development, and organizational change. Some of her previous clients include Sterling Events, Lanikila Health and Wellness, Kinetic Heights, and Atrium Health.

https://orcid.org/0000-0002-1591-1910

## References

Abdelnour, S., Hasselbladh, H., & Kallinikos, J. (2017). Agency and institutions in organization studies. *Organization Studies, 38*, 1775–1792. https://doi.org/10.1177/0170840617708007

Ågerfalk, P. J. (2020). Artificial intelligence as digital agency. *European Journal of Information Systems, 29*(1), 1–8. https://doi.org/10.1080/0960085X.2020.1721947

Bailey, D. E., & Barley, S. R. (2020). Beyond design and use: How scholars should study intelligent technologies. *Information and Organization, 30,* 100286. https://doi.org/10.1016/j.infoandorg.2019.100286

Banks, J., & de Graaf, M. (2020). Toward an agent-agnostic transmission model: Synthesizing anthropocentric and technocentric paradigms in communication. *Human-Machine Communication, 1*, 19–36. https://doi.org/10.30658/hmc.1.2

Banks, S. P., & Riley, P. (1993). Structuration theory as an ontology for communication research. *Communication Yearbook 16*, 167–196. https://doi.org/10.1080/23808985.1993.11678851

Barley, S. (1986). Technology as an occasion for structuring: Evidence from observations of CT scanners and the social order of radiology departments. *Administrative Science Quarterly, 31*, 78–108. https://doi.org/10.2307/2392767

Beane, M. (2019). Shadow learning: Building robotic surgical skill when approved means fail. *Administrative Science Quarterly, 64*, 87–123. https://doi.org/10.1177/0001839217751692

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition, 84*, 317–331. https://doi.org/10.1016/j.patcog.2018.07.023

Boulding, K. E. (1956). General systems theory—The skeleton of science. *Management Science, 2*, 197–208. https://doi.org/10.1287/mnsc.2.3.197

Brynjolfsson, E., Mitchell, T., & Rock, D. (2018, May). What can machines learn, and what does it mean for occupations and the economy? *AEA Papers and Proceedings*, *108*, 43–47.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*, 416–431. https://doi.org/10.1080/21670811.2014.976412

Chan, N. K., & Humphreys, L. (2018). Mediatization of social space and the case of Uber drivers. *Media and Communication, 6*, 29–38. https://doi.org/10.17645/mac.v6i2.1316

Christin, A. (2017). Algorithms in practice: Comparing web journalism and criminal justice. *Big Data & Society*, *4*, 1–14. https://doi.org/10.1177/2053951717718855

Chui, M., & Malhotra, S. (2018). *Notes from the AI frontier: AI adoption advances, but foundational barriers remain*. https://www.mckinsey.com/featured-insights/artificial-intelligence/ai-adoption-advances-but-foundational-barriers-remain

DeSanctis, G., & Poole, M. S. (1994). Capturing the complexity in advanced technology use: Adaptive structuration theory. *Organization Science, 5*, 121–147. https://doi.org/10.1287/orsc.5.2.121

Diktonius, J. (2016, October 17). *Tieto is the first Nordic company to appoint Artificial Intelligence to the leadership team of the new data-driven business unit*. https://web.archive.org/web/20190919205734/https://www.tieto.com/en/newsroom/all-news-and-releases/corporate-news/2016/10/tieto-the-first-nordic-company-to-appoint-artificial-intelligence-to-the-leadership-team-of-the-new-data-driven-business/

Duhigg, C. (2012, February 16). How companies learn your secrets. *New York Times.* Retrieved from https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=1&_r=1&hp

Edwards, R. C. (1981). The social relations of production at the point of production. In Zey-Ferrell, M., & Aiken, M. (Eds.). *Complex organizations: Critical perspectives* (pp. 156–182). Scott, Foresman.

Ekbia, H., & Nardi, B. (2014). Heteromation and its (dis)contents: The invisible division of labor between humans and machines. *First Monday, 19*(6). https://doi.org/10.5210/fm.v19i6.5331

Elliott, A. (2019). *The culture of AI: Everyday life and the digital revolution*. Routledge.

Emirbayer, M., & Mische, A. (1998). What is agency? *American Journal of Sociology, 103*, 962–1023. https://www.jstor.org/stable/10.1086/231294

Faraj, S., Pachidi, S., & Sayegh, K. (2018). Working and organizing in the age of the learning algorithm. *Information and Organization, 28*, 62–70. https://doi.org/10.1016/j.infoandorg.2018.02.005

Foucault, M. (1979). Authorship: What is an author? *Screen*, *20*(1), 13–34. https://doi.org/10.1093/screen/20.1.13

Fraune, M. R. (2020). Our robots, our team: Robot anthropomorphism moderates group effects in human–robot teams. *Frontiers in Psychology, 11*(1275), 1–14. https://doi.org/10.3389/fpsyg.2020.01275

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*, 71–86. https://doi.org/10.30658/hmc.1.5

Gaudiello, I., Zibetti, E., Lefort, S., Chetouani, M., & Ivaldi, S. (2016). Trust as indicator of robot functional and social acceptance. An experimental study on user conformation to iCub answers. *Computers in Human Behavior, 61*, 633–655. https://doi.org/10.1016/j.chb.2016.03.057

Giddens, A. (1979). *Central problems in social theory: Action, structure and contradiction in social analysis.* University of California Press.

Giddens, A. (1984). *The constitution of society.* University of California Press.

Graefe, A., Haim, M., Haarmann, B., & Brosius, H.-B. (2018). Readers' perception of computer-generated news: Credibility, expertise, and readability. *Journalism*, *19*, 595–610. https://doi.org/10.1177/1464884916641269

Gray, M. L., & Suri, S. (2019). *Ghost work: How to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st Century. *Communication +1, 1*(1), Article 1. https://doi.org/10.7275/R5QJ7F7R

Guzman, A. L. (2016). The messages of mute machines: Human-Machine Communication with industrial technologies. *Communication +1, 5*(1), Article 4. https://doi.org/10.7275/R57P8WBW

Guzman, A. L. (Ed.). (2018). *Human-Machine Communication: Rethinking communication, technology, and ourselves.* Peter Lang Publishing, Inc.

Guzman, A. L. (2020). Ontological boundaries between humans and computers and the implications for Human-Machine Communication. *Human-Machine Communication, 1*, 37–54. https://doi.org/10.30658/hmc.1.3

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society, 22*, 70–86. https://doi.org/10.1177/1461444819858691

Howard, L. A., & Geist, P. (1995). The dialectic of control in a merging organization. *Communication Monographs, 62,* 110–131. https://doi.org/10.1080/03637759509376352

IBM Watson Talent. (2021). https://www.ibm.com/services/talent-management?mhsrc=ibm-search_a&mhq=Watson%20Talent

Janlert, L. E., & Stolterman, E. (2017). The meaning of interactivity: Some proposals for definitions and measures. *Human–Computer Interaction, 32*(3), 103–138. https://doi.org/10.1080/07370024.2016.1226139

Kellogg, K. C., Valentine, M. A., & Christin, A. (2020). Algorithms at work: The new contested terrain of control. *Academy of Management Annals*, *14*, 366–410. https://doi.org/10.5465/annals.2018.0174

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems, 25*, 1097–1105. https://doi.org/10.1145/3065386

Latour, B. (1991). Technology is society made durable. In J. Law (Ed.), *A sociology of monsters. Essays on power, technology and domination* (pp. 103–131). Routledge.

Lee, C., Rahafrooz, M., & Lee, O. K. D. (2017). What are the concerns of using a ride-sharing service? An investigation of Uber. In *Proceeding of the 23rd Americas Conference on Information Systems* (AMCIS 2017), 10–12 August, Boston, Massachusetts.

Leonardi, P. M. (2011). When flexible routines meet flexible technologies: Affordance, constraint, and the imbrication of human and material agencies. *MIS Quarterly, 35*, 147–167. https://doi.org/10.2307/23043493

Levy, K. E. C. (2015). The contexts of control: Information, power, and truck-driving work. *The Information Society, 31*, 160–174. https://doi.org/10.1080/01972243.2015.998105

Lewis, S. C., Guzman, A. L., & Schmidt, T. R. (2019). Automation, journalism, and human-machine communication: Rethinking roles and relationships of humans and machines in news. *Digital Journalism, 7,* 409–427. https://doi.org/10.1080/21670811.2019.1577147

Majchrzak, A., Rice, R., Malhotra, A., King, N., & Ba, S. (2000). Technology adaptation: The case of a computer-supported inter-organizational virtual team. *MIS Quarterly, 24*, 569–600. https://doi.org/10.2307/3250948

Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together*. Little, Brown, and Company.

McPhee, R. D., Poole, M. S., & Iverson, J. (2014). Structuration theory. In L. L. Putnam & D. K. Mumby (Eds.), *The SAGE handbook of organizational communication: Advances in theory, research, and methods* (3rd ed., pp. 75–100). Sage Publications.

Moore, A. (2019, June 7). When AI becomes an everyday technology. *Harvard Business Review*. https://hbr.org/2019/06/when-ai-becomes-an-everyday-technology

Neff, G., & Nagy, P. (2016). Automation, algorithms, and politics| Talking to Bots: Symbiotic agency and the case of Tay. *International Journal of Communication, 10*, 4915–4931. https://ijoc.org/index.php/ijoc/article/view/6277

Newlands, G. (2020, online). Algorithmic surveillance in the gig economy: The organization of work through Lefebvrian conceived space. *Organization Studies*. https://doi.org/10.1177/0170840620937900

Noble, D. F. (2011). *Forces of production: A social history of industrial automation*. Transaction Publishers.

O'Neil, C. (2016). *Weapons of math destruction: How big data increases inequality and threatens democracy*. Broadway Books.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. *Organization Science, 3*, 398–427. https://doi.org/10.1287/orsc.3.3.398

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies, 28*, 1435–1448. https://doi.org/10.1177/0170840607081138

Orlikowski, W. J., & Scott, S. V. (2008). Sociomateriality: Challenging the separation of technology, work and organization. *Academy of Management Annals, 2*, 433–474. https://doi.org/10.5465/19416520802211644

Orlikowski, W. J., & Yates, J. (1994). Genre repertoire: The structuring of communicative practices in organizations. *Administrative Science Quarterly, 39*, 541–574. https://doi.org/10.2307/2393771

Perc, M., Ozer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications, 5*, 1–7. https://doi.org/10.1057/s41599-019-0278-x

Pondy, L. R., & Mitroff, I. I. (1979). Beyond open systems models of organizations. In B. M. Staw (Ed.), *Research in Organization Behavior* (vol. 1, pp. 3–39). JAI Press.

Pozzebon, M., & Pinsonneault, A. (2005). Challenges in conducting empirical work using structuration theory: Learning from IT research. *Organization Studies, 26,* 1353–1376. https://doi.org/10.1177/0170840605054621

Raghuram, S., Hill, N. S., Gibbs, J. L., & Maruping, L. M. (2019). Virtual work: Bridging research clusters. *Academy of Management Annals, 13*, 308–341. https://doi.org/10.5465/annals.2017.0020

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

Riley, P. (1983). A structurationist account of political culture. *Administrative Science Quarterly, 28*(3), 414–437. https://doi.org/10.2307/2392250

Rosenblat, A. (2019). *Uberland: How algorithms are rewriting the rules of work.* University of California Press.

Rosenblat, A., & Stark, L. (2016). Algorithmic labor and information asymmetries: A case study of Uber's drivers. *International Journal of Communication, 10*, 3758–3784. https://doi.org/10.2139/ssrn.2686227

Russell, S., & Norvig, P. (2019). *Artificial intelligence: A modern approach* (4th ed.). Pearson Education. https://doi.org/10.1017/S0269888900007724

Sparrow, B., Liu, J., & Wegner, D. M. (2011). Google effects on memory: Cognitive consequences of having information at our fingertips. *Science, 333*, 776–778. https://doi.org/10.1126/science.1207745

Stillman, L., & Stoecker, R. (2005). Structuration, ICTs, and community work. *The Journal of Community Informatics, 1*(3), 83–102.

Sturken, M., Thomas, D., & Ball-Rokeach, S. J. (Eds.). (2004). *Technological visions: The hopes and fears that shape new technologies.* Temple University Press.

Suchman, L. A. (2007). *Human-machine reconfigurations: Plans and situated actions* (2nd ed.) Cambridge University Press.

Sundar, S. S., Bellur, S., Oh, J., Jia, H., & Kim, H.-S. (2016). Theoretical importance of contingency in human-computer interaction: Effects of message interactivity on user engagement. *Communication Research, 43,* 595–625. https://doi.org/10.1177/0093650214534962

Sundararajan, A. (2016). *The sharing economy: The end of employment and the rise of crowd-based capitalism*. The MIT Press. https://doi.org/10.2307/j.ctt1c2cqh3

Tompkins, P. K., & Cheney, G. (1985). Communication and unobtrusive control in contemporary organizations. In R. D. McPhee & P. K. Tompkins (Eds.), *Organizational communication: Traditional themes and new directions* (pp. 179–210). Sage Publications.

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Weber, M. S., & Kosterich, A. (2018). Coding the news: The role of computer code in filtering and distributing news. *Digital Journalism, 6*, 310–329.

West, D. M., & Allen, J. R. (2018, April 24). *How artificial intelligence is transforming the world.* Brookings. https://www.brookings.edu/research/how-artificial-intelligence-is-transforming-the-world/

Wiener, N. (1950). *The human use of human beings: Cybernetics and society.* Houghton Mifflin.

Wood, A. J., Graham, M., Lehdonvirta, V., & Hjorth, I. (2019). Good gig, bad gig: Autonomy and algorithmic control in the global gig economy. *Work, Employment, & Society, 33*, 56–75. https://doi.org/10.1177/0950017018785616

Yan, B., Lewis, K., Figge, P., Hollingshead, A., Alexander, K. S., Kim, Y. J., & Fang, C. (2020). Intelligent machines and teamwork: Help or hindrance? In *Academy of Management Proceedings* (Vol. 2020, No. 1, p. 21962). Briarcliff Manor, NY 10510: Academy of Management.

Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power.* Public Affairs.

HUMAN-MACHINE COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.9

# Artificial Intuition in Tech Journalism on AI: Imagining the Human Subject

**Jacob Johanssen[1] and Xin Wang[2]**

1 Institute of Business, Law and Society, St. Mary's University, Twickenham, London, UK
2 College of Arts and Media, Tongji University, Shanghai, People's Republic of China

## Abstract

Artificial intuition (AI acting intuitively) is one trend in artificial intelligence. This article analyzes how it is discussed by technology journalism on the internet. The journalistic narratives that were analyzed claim that intuition can make AI more efficient, autonomous, and human. Some commentators also write that intuitive AI could execute tasks better than humans themselves ever could (e.g., in digital games); therefore, it could ultimately surpass human intuition. Such views do not pay enough attention to biases as well as transparency and explainability of AI. We contrast the journalistic narratives with philosophical understandings of intuition and a psychoanalytic view of the human. Those perspectives allow for a more complex view that goes beyond the focus on rationality and computational perspectives of tech journalism.

**Keywords:** Artificial Intelligence (AI), artificial intuition, human subject, psychoanalysis, technology journalism

## Introduction

Artificial Intelligence (AI) is an evolving and significant technology. It is increasingly discussed in the media and there are many dedicated popular journalistic websites that cover the latest trends and ideas of AI development (Bory, 2019; Brennen et al., 2018; Goode, 2018; Natale & Ballatore, 2017). Such discussions are often characterized by "magical thinking,"

**CONTACT** Dr. Jacob Johanssen • Senior Lecturer in Communications • Institute of Business, Law and Society • St. Mary's University, Twickenham • London, TW1 4SX • United Kingdom • jacob.johanssen@stmarys.ac.uk

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

hyped up and exaggerated claims, and not necessarily very balanced (Goode, 2018, p. 186). AI seems capable of inspiring popular accounts that show the authors' amazement, enthusiasm for, but also fear of, the technology (Cave et al., 2018).

Rather than focusing on the accuracy, the authors' emotional investment, or implied audience response to AI media coverage as some scholars have done, this article follows a different aim. In focusing on one specific trend of contemporary AI, not AI as a whole, we are interested in analyzing it in relation to the following research question: What kind of view of the human subject is articulated in technology journalism about artificial intuition? This broad question refers to both how artificial intuition is discussed in relation to alleged human characteristics that are inherent to the technology, and what kind of view of the human in general is evident in the examined articles. We pay specific attention to how this phenomenon is being discussed on journalistic websites with a focus on technology (the websites iTech Post, Toolbox, TechXplore, Inverse, Business World, Science Mag, TechCrunch, Science Daily, Venture Beat, KDnuggets, Hackermoon, Inc, and Hadean were part of the sample) which cater for a particular audience, rather than more "mainstream" quality journalistic websites such as those of broadsheet or tabloid newspapers. Brennen et al. (2020) define technology journalism as coverage of "the technology industry" (p. 16) by which they mean reporting on the products of technology businesses, on technology businesses themselves, and on the social effects of technology. Tech journalism often covers emerging and new technologies, such as artificial intelligence, communication technology, bio- or nanotechnology. Brennen et al. (2020) argue that it is an understudied phenomenon and "there is a rich potential for future studies of technology journalism" (p. 15). Artificial intuition is so recent and somewhat niche that it has not been adequately covered by tabloid or broadsheet journalism yet. This article therefore specifically focuses on online tech journalism. While each sampled website may have its own focus, target audience, style, and mission, all can be seen as belonging to the category of tech journalism. It would be beyond the scope of the initial research into the topic to discuss each sampled outlet in more detail. Journalistic content is sometimes written by individuals who are, for example, data scientists or AI experts rather than journalists. It is also questionable if some of the outlets examined in this article employ the same measures of quality control as traditional journalism as authors for some could self-publish their views on the topic (Brennen et al., 2020). Future research could situate the research findings more in relation to the specificities of the different outlets of tech journalism.

The central point of this article is that the journalistic narratives that were analyzed claim that AI can mimic and adopt characteristics of human intuition. The data we discuss show a view of the human subject that articulates that AI can both learn from and that it can be better and more advanced than human intuition. We argue that such views demonstrate the *desire* of many journalists who write about AI to make it human-like. It is argued that intuition can make AI more efficient, autonomous, human, and flexible than it currently is. As we discuss in the last section on AI and games, the ultimate hope for some commentators is that tasks can be executed more efficiently than humans themselves ever could. Therefore, intuitive AI could ultimately surpass human intuition. However, such views do not pay enough attention to questions of the human subject as well as transparency, explainability, and accountability of AI.

## Literature Review

There are many different definitions of AI (see, e.g., Russell & Norvig, 2010; Turner, 2019 for overviews). Generally, artificial intelligence refers to "the study of agents that receive percepts from the environment and perform actions" (Russell & Norvig, 2010, p. viii). Such agents are often machines, such as digital computers. Turner defines AI as the "ability of a non-natural entity to make choices by an evaluative process" (p. 16). Non-natural, for Turner, both refers to human-made as well as machine-made (when AI systems create other AI for example).

While AI technologies have been in development since the 1950s, recent years have seen advancements in the complexity of AI when it comes to emulating cognitive characteristics of humans. On a general level, AI has seen increasing developments in the past 10–15 years in particular thanks to the growth of better hardware, data processing, and increased capacity to store and analyze large datasets (Turner, 2019). The seeming ubiquity of AI has led to both an increase in scholarly discussions in the humanities and social sciences as well as AI being picked up by news media. Given that as a technology AI often seeks to simulate or emulate *human* behavior (such as problem-solving, cognitive abilities, or pattern recognition) and advance such human characteristics, it is unsurprising that it has attracted much attention by journalists and scholars alike.

*Artificial intuition* refers to the ability of AI systems to make intuitive choices or respond intuitively to problems (Crowder & Friess, 2013). It was chosen as an example of current discussions of trends in AI that are picked up by journalistic websites with a focus on technology. Questions around intuitive AI have been gaining in prominence in recent years (e.g., when it comes to discussions about self-driving cars or the ability of AI systems that are used for gaming, such as DeepMind's AlphaGo).

Artificial intuition is compared to, for example, neuromorphic hardware, machine learning, or neural networks, more in its infancy (Crowder & Friess, 2013; Srdanov et al., 2016; Tao & He, 2009). As we show, there is some hype around artificial intuition. This makes it all the more relevant to analyze. It has also been discussed in literature on AI that comes from a computer science or engineering perspective for some time and has gained more attention in recent years (Caudill & Butler, 1990; Crowder & Friess, 2013; Diaz-Hernandez & Gonzalez-Villela, 2015; Dundas & Chik, 2011; Frantz, 2003; Johnny et al., 2019; Srdanov et al., 2016; Tao & He, 2009).

AI acting intuitively may be necessary in certain situations, because decisions can be made quicker than in logical AI. Johnny et al. (2019) define artificial intuition as "the ability of a system to assess a problem context and use pattern recognition or properties from a dataset to choose a course of action or aid the decision process in an automatic manner" (p. 470). Srdanov et al. (2016) propose a trial and error approach which combines logic and randomness in order to reach a specific solution or solve a problem (e.g., making moves in a game). They argue that this provides a way to reach a goal quicker than without trial and error. Diaz-Hernandez and Gonzalez-Villela (2015) suggest that human intuition can be mapped onto artificial intuition through three shared characteristics: inputs, processing, and outputs. In acting intuitively, both users and AI systems unconsciously process inputs, act upon them, and generate outputs or solutions. They tested their model through a robot that picked and placed objects on a surface. Johnny et al. (2019) note that one of the potential benefits of artificial intuition is reducing "the complexity of the set of instructions needed to solve the task" (p. 466); for example, in an algorithm.

From a technical perspective, intuition is seen in the above works as "subconscious pattern recognition" (Frantz, 2003, p. 266) of humans and AI systems. All of the authors discussed above seek to translate human intuition into a technical concept which can be operationalized. We argue that the way intuition is approached in those studies is, perhaps paradoxically, anti-intuitive. Intuition is conceptualized as "scientific—rational, logical" (Frantz, 2003, p. 267). This was often mirrored in the way it is presented in the sample of tech journalism articles.

Such forms of rationalistic accounts of intuition are in contrast to how it is theorized in cultural studies and the humanities more generally. In trying to map intuition onto computer science, and thereby essentially making it code-based, the above accounts lose the complexity of intuition and how it relates to dimensions beyond language and consciousness. For instance, human intuition is not the same as a trial and error approach in computation, as Srdanov et al. (2016) claim. While some computer scientists have acknowledged that artificial intuition will remain limited compared to its human counterpart (Diaz-Hernandez & Gonzalez-Villela, 2015), there is scope to put such scholarship in dialogue with other disciplines.

The term "intuition" has been discussed from many different perspectives. For the purpose of this article we shall largely draw on how intuition has been defined in cultural studies, affect theories, and psychoanalysis (for extensive discussion of the term, see Chudnoff, 2013). Intuition is dependent on an individual human subject and therefore differs for each person. It is commonly understood as a form of knowing that is experienced and felt rather than fully consciously known. It is often associated with terms like *gut feeling*, *hunch*, or *having a sense of something*. Intuition designates a bodily state or experience upon which the individual acts without necessarily having the (empirical) evidence or proof. It is something transitory that is followed or listened to and may then be acted upon without fully understanding why. It refers to a mode of experience within immediacy. It is happening in the now, before being fully rendered conscious and cognitively apprehended by the individual. Intuition is not opposed to, or in contradiction to reason, however. Greg Seigworth (2006) has written that "experience and experiment through the faculty of intuition [. . .] exceeds or overflows the intellect" (p. 118). For Carolyn Pedwell (2019), intuition "is embodied experience prior to, or in excess of, its translation into the parsing categories of representational and analytical thought" (p. 127). Lauren Berlant (2008) has characterized intuition as an acute source of knowledge. It is a form of "affective intelligence" (p. 852) that shows itself in a particular relational sensitivity to the worldly surroundings of humans.

"With intuition, experience is less a discrete place in the time past belonging to a subject, and more an immanent process of relation (beyond inside and outside, beyond subject and object)" (Seigworth, 2006, p. 119). Such conceptualizations of intuition broaden its theoretical frame to include a relational and open perspective.

We were specifically interested in how human subjectivity is discussed in tech journalism's articles on artificial intuition. This article is informed by a psychoanalytic theory of the human subject. This means that the human subject and specifically their cognitive functions go beyond rationality and mechanistic accounts which often link computational facilities to the brain (Turkle, 2005). According to psychoanalysis, humans are shaped by emotional, affective, intuitive, as well as rational actions and feelings (Rosenblatt & Thickstun, 1994).

Psychoanalysis holds that humans are intuitive, complex, inefficient, contradictory, and sometimes flawed beings. Human beings are influenced by unconscious processes which shape their fantasies, desires, and actions. This emphasis of the unconscious includes considerations of intuition (Board, 1958; Rosenblatt & Thickstun, 1994; Szalita-Pemow, 1955). Humans regularly make unconscious decisions and this includes intuition but also goes beyond it. For Sigmund Freud, it is both the intuition of the psychoanalyst as well as that of the patient that enables the free associative form of talking in the consulting room during a therapy session (Freud, 1981). For psychoanalysis, then, intuition refers to ways of knowing and experiencing the world where parts of knowledge "are unconscious or otherwise unavailable for conscious delineation" (Board, 1958, p. 237).

## Artificial Intelligence in the Media

The question how artificial intelligence is discussed in the media and in particular what kind of image of the human, is shown through such narratives and has been researched only by a few scholars (Bory, 2019; Brennen et al., 2018; Goode, 2018; Natale & Ballatore, 2017). Natale and Ballatore (2017) examined early media coverage of AI in technological magazines from the 1950s to the 1970s. They found that the articles they examined created a "myth" (p. 2) of AI inevitably becoming a thinking machine which would be able to simulate the human brain. Bory compared the coverage on IBM's Deep Blue in 1997 and DeepMind's AlphaGo in 2016. He argues that there has been a shift in narratives from suspicion and conflicted views about Deep Blue in the late 1990s toward recent discourses about AlphaGo that frame it as embodying human characteristics such as beauty or trust while simultaneously being distinctly unhuman-like. This leads to a view of AI as complementary to humans and opening up ways of collaboration rather than competition between the two. Bourne (2019) similarly discusses how AI is framed by PR strategists to promote particular understandings of AI as "friendly" and "competitive." This masks problematic aspects of AI such as its deployment for purposes of discrimination, or harassment. Goode (2018) analyzed mainstream science fiction, media events (such as AlphaGo), and futurology discourses around AI and argues that they are often sensationalist and misleading, for example, when it comes to discourses about a coming artificial mind. Yet, such discourses can also serve as useful entry points to initiate more nuanced public debate about AI, as Goode argues.

Brennen et al. (2018) have analyzed how UK media have covered artificial intelligence over 8 months. They show that there can be a bias toward industry sources in media coverage and that sources from academia, government, and wider civil society were in the minority in their sample. AI is often being politicized in media coverage when it comes to ethical questions such as bias, automation, national security, discrimination, and other issues. Such questions are downplayed by the industry in order not to damage its image. They conclude that there is a risk of covering AI in a one-sided manner. Cave et al. (2018) have analyzed how researchers, communicators, policymakers, and different publics talk about AI. Their analyses show that "[p]revalent AI narratives share dominant characteristics: a focus on embodiment; a tendency towards utopian or dystopian extremes; and a lack of diversity in creators, protagonists, and types of AI" (p. 4).

No research has specifically focused on tech journalism's coverage of artificial intuition in relation to human subjectivity. There is, thus, scope to further research on the discussion of particular aspects of AI by technology journalism outlets, particularly in relation to how the human subject is imagined in them.

## Methodology

The research that forms the basis of this article specifically focused on one technical aspect of AI and how it was covered by tech journalism websites. The articles were obtained by performing a keyword search using Google UK (general search) with the keywords "artificial intuition" and "artificial AND intuition" as well as "artificial intui*" (to include variations of the word, such as "intuitive") in May 2019 and September 2020. It needs to be acknowledged that using Google as a sampling procedure can be seen as problematic but is nonetheless used by scholars as a form of data collection (Ballatore, 2015; Wouters & Gerbec, 2006). The researchers' institutions had no access to databases such as Factiva that would specifically include online sources (rather than only newspaper articles). Given the exploratory nature of this study, a broad approach to sampling was taken in order to include a diverse range of websites. Using a newspaper database such as LexisNexis would have restricted the sample, because tech journalism websites are not indexed. Additionally, a keyword search via LexisNexis did not return any results for "artificial intuition" from 2012 onward. Google was thus pursued as the most pragmatic sampling procedure. The search was initially performed using a university computer, based in London (UK). Using Google as a basis for data collection may shape how results are presented because of the location of the search queries, personalized Google profiles, or cookies saved on the computer. In order to replicate the sampling procedure, the same searches were performed using a public computer (without cookies being stored) at the same university. Those procedures were followed in order to obtain "results as close to the product's default results as possible" (Ballatore, 2015, online).

As the phenomenon of this article is relatively new, only articles published from 2016–2019 were selected for inclusion in order to analyze the most recent discussions. Articles from the first four pages of Google (10 results per page) were selected for analysis in order to select the most popular results (Ballatore, 2015). The following sampling procedure was followed: Articles selected for sample inclusion needed to feature a discussion of artificial intuition that was longer than 300 words. Only articles from professional websites and not from personal homepages (private homepages created by individuals) were selected. Articles needed to fall into a broad definition of tech journalism that catered for an audience with an interest in technology and specifically AI (Brennen et al., 2020; Natale & Ballatore, 2017). Only articles that sought to explain artificial intuition were included. A total of 29 articles were sampled (see annex table for an overview of all sampled articles). A small sample was deliberately chosen in order to conduct a qualitative, detailed thematic textual analysis of the data (Mayring, 2000; Saldaña, 2009). A total of seven items are discussed in this article. Qualitative data analysis was conducted as follows: The content was read and then coded into themes using Nvivo, some of which are outlined further in the following sections (Mayring, 2000; Saldaña, 2009). Themes were constructed as summaries of common points made across articles in order to group articles together. The themes were analyzed through a qualitative thematic analysis with a specific focus on how AI technologies were being discussed in relation to human subjectivity. We thus paid particular attention to how AI was described in relation to the human (e.g., through evoking human characteristics, such as the body, brain functioning, cognition, or intuition). Two themes were of a more general nature and dealt with what artificial intuition is and how it relates to human intuition. They were included in this article to introduce readers to an analysis of the topic. A third theme in relation to discussions of practical implementations of artificial intuition in games was selected in order to include journalistic discussions beyond definitions or broader narratives. The tables below provide a summary of the items of this article and their themes and coding rules.

**TABLE 1 Overview of Sampled Items**

| Theme and Summary | Number of Items | Coding Rules |
|---|---|---|
| *#1 Defining Artificial Intuition:* Defined in relation to human intuition | In this article: Four | Article must include general discussions of artificial intuition which define it |
| *#2 Intuition and Rationality:* Discussed as making AI more rational and unbiased | In this article: Two | Article must include discussions of AI as becoming more rational and unbiased when being intuitive |
| *#3 Practical Implications of Artificial Intuition in Games:* Discussed as surpassing human intuition | In this article: Two | Article must include discussions of practical implications of AI for AI-based digital games |

**TABLE 2 Further Information on the Sampled Items**

| Title | Outlet/Author | Length and Theme | URL |
|---|---|---|---|
| Deep Learning, Artificial Intuition and the Quest for AGI | KDNuggets/Carlos Perez | 1,132 words #1 | Link 1 |
| Artificial Intuition | Medium/Kees Groeneveld | 1,739 words #1 | Link 2 |
| Artificial Intuition | Medium/Nell Watson | 677 words #1 | Link 3 |
| Did A.I. just make the leap to being intuitive? | Inc/Thomas Koulopoulos | 483 words #1 and #2 | Link 4 |
| Artificial intuition will supersede artificial intelligence, experts say | NetworkWorld/Patrick Nelson | 561 words #2 | Link 5 |
| Building worlds to grow the artificial mind: The AI petri-dish | Hadean/Rashid Mansoor | 522 words #3 | Link 6 |
| A computer's newfound "intuition" beats world poker champs | CNN Health/Michael Nedelman | 1,378 words #3 | Link 7 |

While a larger sample might have been beneficial, this article focuses on detailed discussion and therefore features a limited amount of data. Future research could devote more scope for further empirical analyses. A more complex sampling procedure could also be followed (e.g., by including tabloid and broadsheet newspapers of specific countries).

## Defining Artificial Intuition

Eight items in the sample discuss artificial intuition in a more general manner. For instance, Groeneveld (2018) makes the argument when writing about Deep Learning and AI that such a type of AI is actually a form of artificial intuition, because we do not fully know how Deep Learning networks come to acquire their knowledge. Frank Pasquale (2015) and others have critically discussed AI through the metaphor of the black box. They argue that, for example, AI-based algorithms deliberately obfuscate their operations from users, competitors, and governments in the interests of competition, surveillance, and profit maximization (see also Burrell, 2016; Cheney-Lippold, 2017; Finn, 2017). The journalistic outlets examined are less critical and instead are more celebratory in tone. This may be because of the wider ways in which technology is often framed by such outlets as something inherently positive and useful for humans, as scholars have pointed out (Bourne, 2019; Brennen et. al., 2020; Goode, 2018; Natale & Ballatore, 2017). This points to a problem of such narratives in which the *appearance* of AI is mistaken for its *essence*: it is argued that AI can make intuitive decisions which are human-like. "They [neural networks] are able to creatively fill in gaps and make intuitive leaps to make an appropriate response to a given situation" (Watson, 2017, online), writes one commentator, for example. An article on the tech website KDnuggets similarly suggests:

> Deep Learning systems exhibit behavior that appears biological despite not being based on biological material. It so happens that humanity has luckily stumbled upon Artificial Intuition in the form of Deep Learning. (Perez, 2017, online)

The equation between humans and AI based on the appearance or particular exhibition of certain processes is simplistic. Intuition is narrowly characterized as something computational:

> First off, intuition is just a label we use for a correct decision that's based on incomplete knowledge. We're okay if people are intuitive, in fact we elevate and admire them for it, but we're unsettled by the prospect of a machine making a decision that involves intuition, ambiguity, or less than complete data. But what if our gut is nothing more than a bunch of variables that we're not consciously aware of? (Koulopoulos, 2017, online)

Such a computational view of the human subject is quite different to how human intuition is seen in the humanities and specifically in psychoanalysis. While intuition is generally seen as a positive characteristic of humans, a psychoanalytic view complicates this. For psychoanalysis, intuition is an unconscious process (Jung, 1977). The psychoanalyst

C. G. Jung (1977) defined intuition as “perception via the unconscious” (p. 306) where one may be perceiving something without conscious awareness. Broadly speaking, one of the goals of psychoanalytic therapy is for individuals to understand better why they acted impulsively, intuitively, and emotionally in particular situations. It understands intuition as being shaped by an individual’s biography and their unconscious fantasies and actions. The patient enters into therapy as they suffer because of past or present conflicts, experiences, or relations. They may act intuitively and subsequently understand their intuition better through talking to a psychoanalyst. Unconscious processes can never be rendered completely conscious and some core of the unconscious will always remain (Freud, 1981; Jung, 1977). We evoke the clinical setting at this point because it illustrates the complexity of intuition that AI does not have. In that sense, AI would not act intuitively in a certain situation as a human might *because* of a past trauma or specific unconscious experience. A psychoanalytic view of intuition and human subjectivity helps to contrast it with the very different understanding of artificial intuition that is put forward in the above narratives. While artificial intuition may appear human-like, it cannot reach the complexity of human intuition as it is unique to each individual. Deploying intuition as a notion or model for AI is therefore problematic.

Based on the definitions of intuition in some of the journalistic outlets, commentators misrecognize intuition as a technicality that can be added to AI in order to make it more flexible, dynamic, and autonomous. As discussed earlier, for cultural theorists, intuition is about an immediate affective and sensorial engagement with the world (Berlant, 2008; Pedwell, 2019; Seigworth, 2006). For commentators on AI, it is portrayed as a *technicality* that should be built. The intuitiveness of AI is code-based. AI cannot sense intuitively in the same way a human can. Intuition is something that would be programmed into AI. It would be defined through language (code) what and how AI can act intuitively. This shows the limits of true intuition, because for cultural studies scholars the term refers to sensual, affective processes that are in tension with or beyond the discursive. Furthermore, humans cannot be taught to be intuitive; they cannot follow a set of pre-defined rules in order to perfect a sense of intuition. Neither could AI. Psychoanalysis understands intuition as associative. Rather than seeing intuitions as forms of unconscious pattern recognition, as some computer scientists do, intuitions can be regarded as “apprehensions for which one has no conscious reason” (Perkins, 1976, p. 120).

What is at stake here, then, is the deployment of intuition with good intentions by journalists (following technical developments in AI) but its underlying technicality may ultimately harm both humans and AI. Artificial intuition becomes just another form of rationality with little connection to how it is understood in other disciplines. We unpack this point further in the following pages.

## Intuition, Unbiased Rationality, and Transparency

Artificial intuition is in all of the sampled articles regarded as something that can make AI more effective and which would ultimately require less human supervision or control. A journalist writes about the development of a planning algorithm at MIT: “The school recently said it now knows how to include human intuition in a machine algorithm. That’s a big deal" (Nelson, 2017, online). Such sentences suggest that intuition can be coded into AI. An article on Inc.com similarly states:

> AI is actually very well suited to making those sorts of highly intuitive decisions. Since it's not conscious, it has no bias as to what it observes and therefore it's aware of everything that influences a particular decision. (Koulopoulos, 2017, online)

The above author argues that AI is unbiased in its observation of its surroundings. However, AI is far from being unbiased. AI-powered systems have been built by humans, with inherent biases, in the first place and those are often (un)consciously built into the technology (Cheney-Lippold, 2017; Noble, 2018; West et al., 2019). Such narratives also leave out sociocultural aspects of humans and AI; for example, biases of AI systems against certain individuals. Such biases mirror wider social inequalities and stereotypes as they are (un)consciously coded into AI. The question if, for example, a neural network can make a decision based on a hunch is significant, because it can potentially grant even more power to such systems. Researchers have pointed to the ineffectiveness and racialized bias of various AI systems (West et al., 2019). Such ethical questions were largely absent from the items in the sample.

The issue of intuition also relates to wider debates concerning transparency and explainability in AI (Burrell, 2016; Felzmann et al., 2019). AI systems log and store their own decisions and processes, so that humans can scrutinize and troubleshoot them. In that sense, AI is always *conscious* (aware) of its own processes. However, it is often not possible to explain to lay people how particular decisions have been made by AI, because they are too complex. Intuition would only add to this, because something partly unexplainable through discourse would be needed to be explained by AI (or humans) in order for it to be accountable. This complicates the notion of "retrospective transparency" of AI. Felzmann et al. (2019) define retrospective transparency as a process that "reveals for a specific case how and why a certain decision was reached, describing the data processing step by step" (p. 2). This means that individuals should be able to understand the different factors that led to particular decisions of an AI (both input- and output-related). However, as Felzmann et al. point out, notions of transparency, as they show themselves in practices of informed consent, for example, often assume a rational and fully autonomous individual (Johanssen, 2019). Yet, there are limits to human rationality, as we have also discussed via our psychoanalytic considerations of subjectivity. Individuals may not always be able or willing to understand complex technological processes (Kemper & Kolkman, 2018). Expectations around transparency also differ according to the specifics of a particular AI (e.g., a social media algorithm or a self-driving car). Users may still trust technology even if they cannot be fully informed about its use (Heald, 2006). However, if the "willingness to be accountable" is seen as "a core indicator of trustworthiness" of AI (Felzmann et al., 2019, p. 10), it remains questionable how such a willingness could be achieved in the case of artificial intuition.

Such questions further complicate ethical considerations of artificial intuition. There is a danger in advocating intuitive AI because intuition is by definition difficult, if not impossible, to explain and account for. If AI acted truly intuitively, this could serve as a justification for being intransparent and opaque, because particular intuitive actions could not be explained. It would further the perception of AI as a black box (Pasquale, 2015). Intuition without accountability and transparency would be dangerous. This would also lead to declining trust in AI (Felzmann et al., 2019).

Going beyond the technical aspects of artificial intuition, how the term is (mis)used in tech outlets shows a particular understanding of human subjectivity. AI is not only anthropomorphized and made more human through the advocacy for intuition, at the same time, humans are made more machinic, algorithmic, and technical than they really are in the data we examine. This is done through formulations such as "intuition is just a label we use for a correct decision that's based on incomplete knowledge," our gut being "a bunch of variables" that we are "not consciously aware of" (Koulopoulos, 2017, online). The human subject is regarded in a one-dimensional, functionalist way. Humans and AI function in the same ways and are almost interchangeable. Such equations do not do justice to the complexity of human subjectivity. The human mind is more contradictory and messier than a computer (Turkle, 2005).

Such narratives are similar to the ones of the computer scientists we discussed earlier which suggest that artificial intuition is something that can be coded based on human intuition. Such an argument is too simplistic and while it may speak to the desire of mapping human characteristics onto AI, there is a risk in understanding the human being only in terms that come from computer science. Human subjectivity, in particular, is not only grounded in rationality and language.

## Applications of Artificial Intuition: Games

But what does it really mean for AI to act intuitively? How is intuition understood and framed in such commentaries in relation to specific applications or forms of usage of AI? The last theme we discuss here is the specific application of artificial intuition to digital games. Alphabet's AlphaGo and AlphaGo Zero have recently made headlines for their ability to defeat professional players of the game Go:

> AlphaGo Zero often optimises for better board position with virtually utter disregard for piece value. Interestingly, this stems from its ability to teach itself rather than observe human play and thus avoid contamination from human biases. [. . .] It relies more strongly on heuristics or intuition, needing to look at fewer board positions to arrive at clever strategies. (Mansoor, 2018, online)

Another article discusses AI used for playing poker against humans:

> To form DeepStack's so-called intuition, Bowling and his team ran millions of test games against the AI. [. . .] [P]layers may bluff to hide bad cards, but those are situations DeepStack has already accounted for through math. (Nedelman, 2017, online)

While the need for intuitive AI systems may be comprehensible, the question remains if there is not a contradiction in terms present here. While human intuition is based on experience and the accumulation of enough *data* to act intuitively in a given situation, how and why the data or experiences were drawn upon when someone acts intuitively is far more fuzzy and complex than the above narratives portray. AlphaGo and DeepStack (which are heavily based on neural networks) may have been able to act more intuitively and teach themselves new strategies, but that is hardly the same as human intuition as such. In fact, as Paolo Bory (2019) discusses, the move that led to the defeat of the world champion by the original AlphaGo (the previous version of AlphaGo Zero) was considered as creative, even beautiful by commentators. It was also seen as very un-humanlike. "The day after [the match], DeepMind revealed that AlphaGo decided to play that move for this very reason since the possibility that a human player would play that move was 1:10,000" (Bory, 2019, p. 10). This explanation of an AI "deciding" to perform a particular move, is rooted in logic rather than intuition. A calculation of probabilities led to a particular move that led to the defeat of the human player. The other two quotations about AlphaGo Zero and DeepStack reproduced earlier similarly describe intuition as logical and math-based. Artificial intuition is seen as something that can surpass human intuition in advancing a form of intuition that is unlike those of humans. However close to intuition the moves by AlphaGo Zero or DeepStack may have been, the data that they trained themselves with were still accessible to them. Both systems *consciously* learned from data.

The narratives discussed in this section point to a different understanding of artificial intuition to the one we have discussed in the previous sections of this article. Whereas many commentators see artificial intuition as something that can be based on human intuition, the discussions of its application in game AIs frame it as something that can actually advance, if not surpass, human intuition as a form of un-humanlikeness. Such narratives suggest a desire for advancing human subjectivity through AI rather than merely mimicking it. Human beings are implicitly seen as deficient or suboptimal in such narratives and it is AI that can advance or assist them. In their work on popular and scientific definitions of social robots, Sarrica et al. (2019) similarly note that "popular online definitions tend to frame social robots as fully autonomous agents, while the scientific literature underscores the fact that they are purpose-built artefacts, entities that execute specific tasks and are only functionally autonomous" (p. 16).

Such narratives are in line with research on media coverage of AI discussed earlier which cover the subject in a future-oriented way that is often somewhat similar to describing science-fiction scenarios, and unlike how scientific discourses would operate (Bourne, 2019; Natale & Ballatore, 2017). The particular journalistic discourses on AI we have presented also need to be seen in relation to the websites they are published in; tech journalism is a particular genre of journalism that Hanusch (2012) links to lifestyle journalism, a genre he defines as providing information to audiences "often in entertaining ways, about goods and services they can use in their daily lives" (p. 4). Brennen et. al. (2020) note that for tech journalists, "story ideas come from technology 'insiders' in the form of company announcements, blogs, or personal Twitter accounts" (p. 15). There is thus a danger of being influenced by industry PR (see also Bourne, 2019). The narratives we have analyzed in this article often read like exaggerated claims of the AI industry in relation to what artificial intuition could do in the future (Sarrica et al., 2019).

While the desire to advance human subjectivity through AI may be comprehensible, it nonetheless presents ethical challenges in relation to transparency. It also presents practical challenges in relation to the amount of control over AI that humans would retain should they be *outsmarted* by intuitive systems. It is beyond our ability to assess the technical probability of artificial intuition, but how it is portrayed and envisioned in the sampled articles is problematic insofar as important ethical and philosophical notions are not considered.

## Conclusion

It was the goal of this article to analyze how artificial intuition is discussed in tech journalism. We were particularly interested to explore what such narratives reveal about the relationship between humans and AI, as well as what kind of understanding of the human subject is implicitly and explicitly shown in the data. One limitation of this study is its exploratory nature and, therefore, small data set that was examined. Future research should include a larger sample in order to discuss a wider variety of sources. The specific styles, orientations, and target audiences of different outlets that focus on technology journalism should also be examined in more detail. Future research could also include tabloid and broadsheet newspapers of specific countries, for example. Comparisons between tech journalism and more mainstream forms of journalism could then be made.

While we cannot adequately assess the technological feasibility of the phenomenon discussed in this article, we argue that it points to particular desires in how AI technology should progress. In a sense, AI seeks to adopt human characteristics both ontologically and epistemologically. Artificial intuition is an example of making AI appear more efficient, autonomous, human, and flexible than it currently is.

There is a particular understanding, or worldview, of the human subject present in the data we have examined. An understanding that sees the human subject as rational and as being a blueprint, ready to be exploited by AI technology so that humans may be enhanced or, as some accounts argue, ultimately surpassed. Something as complex and difficult to define as intuition is rendered into a mechanized concept for the sake of framing AI in particular ways. We have contrasted such views of the human subject with a psychoanalytic prism which conceptualizes the human as irrational, intuitive, emotional, and as being shaped by trauma and conflict.

The hype around artificial intuition may also have been created by the industry because it not only makes AI appear more human, but also more empathic and likeable (Bory, 2019; McStay, 2018). This might be done deliberately in order to mask transparency, explainability, and accountability of AI. As we have discussed, such a scenario raises crucial ethical questions about the transparency of AI.

Intuition, then, is not the best term to make very valid arguments about the current rigidness or limitations of AI. AI needs to become more *flexible*—rather than intuitive—and this should happen in an ethical and transparent manner.

## Author Biographies

**Jacob Johanssen** (PhD) is Senior Lecturer in Communications, St. Mary's University (London, UK). He is the author of *Psychoanalysis and Digital Culture: Audiences, Social Media, and Big Data* (Routledge, 2019). His research interests include psychoanalysis and digital media, emerging technologies and artificial intelligence, psychosocial studies, and critical theory.

https://orcid.org/0000-0002-8682-1648

**Xin Wang** (PhD) is Tenure-Track Professor of Media and Cultural Studies, College of Arts and Media, Tongji University (Shanghai, People's Republic of China). Her research interests are Media and Cultural Studies, intercultural communication, critical theory and media philosophy.

## References

Ballatore, A. (2015). Google chemtrails: A methodology to analyze topic representation in search engine results. *First Monday*, *20*(7).

Berlant, L. (2008). Intuitionists: History and the affective event. *American Literary History*, *20*(4), 845–860. https://doi.org/10.1093/alh/ajn039

Board, R. (1958). Intuition in the methodology of psychoanalysis. *Psychiatry*, *21*(3), 233–239. https://doi.org/10.1080/00332747.1958.11023132

Bory, P. (2019). Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo. *Convergence*. https://doi.org/10.1177/1354856519829679

Bourne, C. (2019). AI cheerleaders: Public relations, neoliberalism and artificial intelligence. *Public Relations Inquiry*, *8*(2), 109–125. https://doi.org/10.1177/2046147X19835250

Brennen, S., Howard, P. N., & Kleis Nielsen, R. (2018). An industry-led debate: How UK media cover artificial intelligence. Report. https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-12/Brennen_UK_Media_Coverage_of_AI_FINAL.pdf

Brennen, S., Howard, P. N., & Kleis Nielsen, R. (2020). Balancing product reviews, traffic targets, and industry criticism: UK technology journalism in practice. *Journalism Practice*. https://doi.org/10.1080/17512786.2020.1783567

Burrell, J. (2016). How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data & Society*, *3*(1), 1–12. https://doi.org/10.1177/2053951715622512

Caudill, M., & Butler, C. (1990). *Naturally intelligent systems*. MIT Press.

Cave, S., Craig, C., Dihal, K. S., Dillon, S., Montgomery, J., Singler, B., & Taylor, L. (2018). Portrayals and perceptions of AI and why they matter. London: The Royal Society. https://www.repository.cam.ac.uk/handle/1810/287193

Cheney-Lippold, J. (2017). *We are data: Algorithms and the making of our digital selves*. NYU Press.

Chudnoff, E. (2013). *Intuition*. Oxford University Press.

Crowder, J., & Friess, S. (2013). Artificial psychology: The psychology of AI. *Proceedings of the 3rd Annual International Multi-Conference on Informatics and Cybernetics*. Orlando, FL.

Diaz-Hernandez, O., & Gonzalez-Villela, V. J. (2015). Analysis of human intuition towards artificial intuition synthesis for robotics. *Mechatronics and Applications: An International Journal (MECHATROJ)*, *1*(1), 23–39. https://doi.org/10.2139/ssrn.3427647

Dundas, J., & Chik, D. (2011): Ibsead: - A self-evolving self-obsessed learning algorithm for machine learning. *International Journal of Computer Science & Emerging Technologies*, *1*(4), 74–79. https://arxiv.org/ftp/arxiv/papers/1106/1106.6186.pdf

Felzmann, H., Fosch Villaronga, E., Lutz, C., & Tamo-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, *6*(1), 1–14. https://doi.org/10.1177/2053951719860542

Finn, E. (2017). *What algorithms want: Imagination in the age of computing*. MIT Press.

Frantz, R. (2003). Herbert Simon. Artificial intelligence as a framework for understanding intuition. *Journal of Economic Psychology, 24*, 265–277. https://doi.org/10.1016/S0167-4870(02)00207-6

Freud, S. (1981). Five lectures on psycho-analysis. *The Standard Edition of the Complete Psychological Works of Sigmund Freud. Volume XI. Five Lectures on Psycho-Analysis, Leonardo Da Vinci and Other Works*. The Hogarth Press and the Institute of Psycho-Analysis.

Goode, L. (2018). Life, but not as we know it: AI and the popular imagination. *Culture Unbound: Journal of Current Cultural Research*, *10*(2), 185–207. http://www.cultureunbound.ep.liu.se/v10/a14/cu18v10a14.pdf

Groeneveld, K. (2018). Artificial Intuition. *Medium*. https://medium.com/@casegreenfield/artificial-intuition-784963f989be

Hanusch, F. (2012). Broadening the focus: The case for lifestyle journalism as a field of scholarly inquiry. *Journalism Practice*, *6*(1), 2–11. https://doi.org/10.1080/17512786.2011.622895

Heald, D. A. (2006). Transparency as an instrumental value. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance? Proceedings of the British Academy* (vol. 135, pp. 59–73). Oxford University Press.

Johanssen, J. (2019). *Psychoanalysis and digital culture: Audiences, social media, and big data*. Routledge.

Johnny, O., Trovati, M., & Ray, J. (2019). Toward a computational model of artificial intuition and decision making. In L. Barolli, H. Nishino, H. Miwa (Eds.), Advances in Intelligent Networking and Collaborative Systems. The 11th International Conference on Intelligent Networking and Collaborative Systems (INCoS-2019). Springer Nature, 463–472.

Jung, C. G. (1977). *C.G. Jung speaking: Interviews and encounters*. Princeton University Press.

Kemper, J., & Kolkman, D. (2018). Transparent to whom? No algorithmic accountability without a critical audience. *Information, Communication & Society*, *22*(14), 2081–2096. https://doi.org/10.1080/1369118X.2018.1477967

Koulopoulos, T. (2017). Did A.I. just make the leap to being intuitive? *Inc.com*. https://www.inc.com/thomas-koulopoulos/did-ai-just-make-leap-to-being-intuitive.html

Mansoor, R. (2018). Building worlds to grow the artificial mind: The AI petri-dish. *Hadean*.

Mayring, P. (2000). Qualitative content analysis. *Forum Qualitative Social Research*, *1*(2), http://www.qualitative-research.net/index.php/fqs/article/view/1089/2385

McStay, A. (2018). *Emotional AI: The rise of empathic media*. Sage.

Natale, S., & Ballatore, A. (2017). Imagining the thinking machine: Technological myths and the rise of artificial intelligence. *Convergence*, *26*(1), 3–18. https://doi.org/10.1177/1354856517715164

Nedelman, M. (2017). A computer's newfound 'intuition' beats world poker champs. *CNN Health*. https://edition.cnn.com/2017/03/02/health/artificial-intelligence-poker-intuition-study/index.html

Nelson, P. (2017). Artificial intuition will supersede artificial intelligence, experts say. *Network World*. https://www.networkworld.com/article/3168320/artificial-intuition-will-supersede-artificial-intelligence-experts-say.html

Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. New York University Press.

Pasquale, F. (2015). *The black box society. The secret algorithms that control money and information*. Harvard University Press.

Pedwell, C. (2019). Digital tendencies: Intuition, algorithmic thought and new social movements. *Culture, Theory and Critique*, *60*(2), 123–138. https://doi.org/10.1080/14735784.2019.1579658

Perez, C. (2017). Deep learning, artificial intuition and the quest for AGI. KDnuggets. https://www.kdnuggets.com/2017/02/deep-learning-artificial-intelligence-quest-agi.html

Perkins, D. N. (1976). The limits of intuition. *Leonardo*, *10*(2), 119–125.

Rosenblatt, A. D., & Thickstun, J. T. (1994). Intuition and consciousness. *The Psychoanalytic Quarterly*, *63*(4), 696–714. https://doi.org/10.1080/21674086.1994.11927433

Russell, S., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed.). Pearson.

Saldaña, J. (2009). *The coding manual for qualitative researchers*. Sage.

Sarrica, M., Brondi, S., & Fortunati, L. (2019). How many facets does a "social robot" have? A review of scientific and popular definitions online. *Information Technology & People, 33*(1), 1–21. https://doi.org/10.1108/ITP-04-2018-0203

Seigworth, G. J. (2006). Cultural studies and Gilles Deleuze. In G. Hall and C. Birchall (Eds.), *New Cultural Studies: Adventures in Theory* (pp. 107–126). Edinburgh University Press.

Srdanov, A., Kovačević, N. R., Vasić, S., & Milovanović, D. (2016). Emulation of artificial intuition using random choice and logic. In *2016 13th Symposium on Neural Networks and Applications (NEUREL)* (pp. 1-6). IEEE. https://doi.org/10.1109/NEUREL.2016.7800114

Szalita-Pemow, A. B. (1955). The "intuitive process" and its relation to work with schizophrenics. *Journal of the American Psychoanalysis Association, 3*(1), 7–18. https://doi.org/10.1177/000306515500300102

Tao, W., & He, P. (2009). Intuitive learning and artificial intuition networks. *2009 Second International Conference on Education Technology and Training* (pp. 297–300). IEEE. https://ieeexplore.ieee.org/abstract/document/5381516

Turkle, S. (2005). *The second self: Computers and the human spirit* (Twentieth anniversary ed.). MIT Press.

Turner, J. (2019). *Robot rules: Regulating artificial intelligence*. Palgrave Macmillan.

Watson, N. (2017). Artificial intuition. The limitations (and ridiculous power) of ANN creativity. *Medium*. https://medium.com/@nellwatson/artificial-intuition-928c3a43ec2a

West, S. M., Whittaker, M., and Crawford, K. (2019). *Discriminating systems: Gender, race and power in AI*. AI Now Institute. https://ainowinstitute.org/discriminatingsystems.html

Wouters, P., & Gerbec, D. (2006). Interactive internet? Studying mediated interaction with publicly available search engines. *Journal of Computer-Mediated Communication, 8*(4). https://doi.org/10.1111/j.1083-6101.2003.tb00221.x

**Annex Table: Complete List of Sampled Articles**

| Title | Author/Outlet | Theme(s) | URL |
|---|---|---|---|
| Artificial Intuition | Kees Groeneveld/ Medium | Defining Artificial Intuition | Link 1 |
| Artificial Intuition | Medium/Nell Watson | Defining Artificial Intuition | Link 2 |
| Next-Generation AI Could Develop Its Own "Human Intuition" | Dyani Sabin/ Inverse | Defining Artificial Intuition; Future; AI and Humans | Link 3 |
| Deep Learning, Artificial Intuition and the Quest for AGI | Carlos E. Perez/ KDnuggets | Defining Artificial Intuition | Link 4 |
| Tech Experts Predict Artificial Intuition Will Be the New Trend In Artificial Intelligence | Victor Thomson/ itechpost | Defining Artificial Intuition; Future | Link 5 |
| How Artificial Intelligence Will Take Over the Supermarket Produce Aisles | Joseph Byrum/ TechCrunch | Defining Artificial Intuition; Business | Link 6 |
| Artificial Intuition Wants to Guide Business Decisions. Can It Improve on "Going With Your Gut"? | Alyssa Newcomb/ Fortune | Business | Link 7 |
| Meet Velas: When Artificial Intuition Boosts Blockchain Capabilities | Alex Dovbnya/ UToday | Business; Technology | Link 8 |
| Node Launches Artificial Intuition Platform to Predict Business Outcomes | No author/BW BusinessWorld | Business | Link 9 |
| Blockchain 3.0—Taking DLT to the Next Level With AI-Integration | Hackermoon/David Cullinan | Business; Technology | Link 10 |
| Next-Gen Software Aims to Lift Personal Computer Into Aware State | Nancy Cohen/ TechXplore | Technology | Link 11 |
| How Do You Code Intuition Into AI? | No author/Loop54 | Technology; Intuition and Rationality | Link 12 |
| Helping Machines Understand Laws of Physics for Intuition on How Objects Should Behave | Rob Matheson/ SciTechDaily | Technology | Link 13 |
| Artificial Intuition and Reinforcement Learning, the Next Steps in Machine Learning | Vince Tabora/ BecomingHumanAI | Technology; Future | Link 14 |
| From Intellect to Intuition: What Happens When We Humanise AI Systems? | Nick Ismail/ Information Age | Defining Artificial Intuition; Future | Link 15 |
| Artificial Intelligence Goes Deep to Beat Humans at Poker | Tonya Riley/ Science Mag | Games | Link 16 |

| Title | Author/Outlet | Theme(s) | URL |
|---|---|---|---|
| A Computer's Newfound "Intuition" Beats World Poker Champs | Michael Nedelman/ CNN Health | Games | Link 17 |
| Building Worlds to Grow the Artificial Mind: The AI Petri-Dish | Hadean/Rashid Mansoor | Games | Link 18 |
| Scientists Teach Computers Fear—To Make Them Better Drivers | Patrick Watson/ Science Mag | Feelings | Link 19 |
| AI: Cars With "Gut Feelings" | Arnold Schlegel/ ZF | Feelings | Link 20 |
| Automating the Future of Design | Jim Bull/Forbes | Feelings; Business | Link 21 |
| MIT Incorporates Human Intuition in Artificial Intelligence to Help Computers Plan Better | No author/ Firstpost Tech2 | AI and Humans | Link 22 |
| Can Artificial Intelligence Replace Human Intuition? | Upasana Bhattacharjee/ Qrius | AI and Humans | Link 23 |
| AI and Human Intuition Go Hand in Hand | Yue Wu/Venture Beat | AI and Humans | Link 24 |
| Modular Deep Learning Could Be the Penultimate Step to Consciousness | Carlos E. Perez/ Medium | AI and Humans | Link 25 |
| Are You Intuitive? Challenge My Machine! | John David Martin/ Hackermoon | AI and Humans; Technology | Link 26 |
| Artificial Intuition Will Supersede Artificial Intelligence, Experts Say | Patrick Watson/ NetworkWorld | AI and Humans; Intuition and Rationality; Future | Link 27 |
| 10 Experts on the Future of Artificial Intelligence | Anirudh V. K./ Toolbox | Future | Link 28 |
| Did A.I. Just Make the Leap to Being Intuitive? | Inc/Thomas Koulopoulos | Defining Artificial Intuition; Intuition and Rationality; Future; Technology | Link 29 |

HUMAN-MACHINE COMMUNICATION

**Human-Machine Communication**
Volume 2, 2021
https://doi.org/10.30658/hmc.2.10

# Automation Anxieties: Perceptions About Technological Automation and the Future of Pharmacy Work

**Cameron W. Piercy[1] and Angela N. Gist-Mackey[1]**

**1** Communication Studies Department, University of Kansas, Lawrence, KS, USA

## Abstract

This study uses a sample of pharmacists and pharmacy technicians (*N* = 240) who differ in skill, education, and income to replicate and extend past findings about socioeconomic disparities in the perceptions of automation. Specifically, this study applies the skills-biased technical change hypothesis, an economic theory that low-skill jobs are the most likely to be affected by increased automation (Acemoglu & Restrepo, 2019), to the mental models of pharmacy workers. We formalize the hypothesis that anxiety about automation leads to perceptions that jobs will change in the future and automation will increase. We also posit anxiety about overpayment related to these outcomes. Results largely support the skills-biased hypothesis as a mental model shared by pharmacy workers regardless of position, with few effects for overpayment anxiety.

**Keywords:** automation, future of work, mental models, pharmacy, structural equation modeling

## Introduction

Human-technology collaboration is ubiquitous in modern organizational contexts. Today's workers rely heavily on technology to remain connected, to schedule meetings, coordinate tasks, and collaborate with others (Colbert et al., 2016). Certain industries are embracing increased automation as part of everyday work. For instance, the pharmacy sector has

**CONTACT** Cameron W. Piercy • Communication Studies Department • University of Kansas • 1440 Jayhawk Blvd. Room 102 • Lawrence, KS 66044 • cpiercy@ku.edu

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

embraced automation in order to improve efficiency and reduce errors in the often-tedious task of filling prescriptions. According to Gebhart (2019), "dispensing robots free up pharmacists and technicians for more profitable clinical services that require human judgment" (para. 1). There are many reasons why industries might favor automation, computers, and artificial intelligence over human labor. Capable machines can complete repetitive and dangerous work, they do not strike (Taipale et al., 2015), show up late, need time off, and generally pay for themselves (West, 2018). Further, while the average factory worker costs more than $36/hour, robots costs as little as $4/hour to operate (Wike & Stokes, 2018). Despite the organizational benefits, employees often view technology as potentially job threatening.

Complex technologies are fundamentally changing the scope and structure of human work. As Edwards et al. (2019) contend, "current trends suggest a near future characterized by more common, normalized, and sustained interactions between people and social robots" (p. 312). Advanced technological machines will continue to reshape the labor market by changing jobs, replacing positions, and generally altering the nature of work (Acemoglu & Restrepo, 2019). Thus, researchers must ask: How do employees perceive the future of human-technology collaboration?

This study seeks to test how pharmacy workers conceive of automation and computerization at work. Specifically, we test the *skills-biased technical change hypothesis (SBTC hypothesis)*, the prediction that low-skill work is the most likely to change or be replaced as computerization and automation increases (Berman et al., 1998; Frey & Osborne, 2017), in a situated pharmacy context where education, income, and skill vary according to role (i.e., pharmacists and pharmacy technicians). This study builds on past evidence to elucidate the mental models of both fear of automation and perceptions that one's work will change or be replaced by machines in the future.

*Mental models,* the "dynamic cognitive frameworks representing spatial, systemic, causal, or situational phenomena" (Banks, 2020, p. 2), serve as frames for interpreting the world. Preconceptions about machines serve as inflection points for decision, choice, and action with regard to both novel and well-known stimuli (Spence et al., 2018). As Banks (2020) succinctly explains, "mental models are sets of ideas for what the world is, how it works, how it unfolds, or what happens in it, and these ideas are used to describe, explain, and predict events or things in the world" (p. 2). Mental models are useful for considering machine communication partners because they encompass social norms, orientations toward machines, and other considerations of how technology has, might, or ought to be used (Mantovani, 1996). Because they are based on a social conception of reality, mental models are connected to and often precede communication behaviors (Fairhurst, 2010). In short, mental models shape and are shaped by communication.

We explore mental models about automation in a sample of pharmacists and pharmacy technicians who, despite working side-by-side, have highly discrepant education, income, and skill-requirements (Wheeler et al., 2019). This sample is ideal because pharmacies have long embraced automation practices (Hynniman & Lamy, 1967; Madden & Dreyfus, 1968), and pharmacies represent an environment marked by high demand for efficiency, role clarity, collaboration, and increasingly, human-machine communication (HMC; Albanese et al., 2010). Further, the pharmacy sector is a compelling context because this sector is increasingly adopting automation to reduce errors, improve productivity, and yield better patient care (Barrett et al., 2012). This study reveals how perceptions of machines manifests in everyday work for pharmacy workers. Below we review relevant literature, introduce research questions and hypotheses, and report findings. We conclude with a discussion and recommendations based on findings from this unique sample regarding HMC and focusing on the role of mental models, feelings of anxiety, and perceptions of the future of work.

## The Skill-Biased Technical Change (SBTC) Hypothesis

Economists do not agree about how automation will alter the future of work. The predominant hypothesis is the *skill-biased technical change hypothesis* (*SBTC hypothesis;* Berman et al., 1998; Frey & Osborne, 2017). The SBTC hypothesis posits that due to automation lower skill jobs will be the most susceptible to change, more likely to be eliminated, and pay inequality will increase in the future. Alternatively, some studies also frame the hypothesis as an increased demand for high-skill jobs (Autor, 2015). For example, Frey and Osborne forecasted that half of occupations are susceptible to automation, based on the complexity of job descriptions.

Though some have argued the SBTC hypothesis is wrong (for discussion, see Frank et al., 2019), the basic premise, that deskilled positions involving simple and repetitive tasks will be increasingly automated, generally holds (Brynjolfsson & Mitchell, 2017). However, there is a "lack of empirically informed models of key microlevel processes" including how people interact with, adapt to, and conceptualize automation (Frank et al., 2019, p. 6531). Our study fills this gap by exploring how American pharmacy workers feel about current and impending automation. We offer insights into how technology might be adopted/avoided by the contemporary labor force.

Historically, the adoption of new technologies in face-to-face relationships has garnered widespread skepticism, which Baym (2015) calls moral panic. From the printing press to the internet, humans tend to both assume and internalize the negative effects of new technologies on relational outcomes—despite a breadth of evidence suggesting otherwise (Green & Clark, 2015). This is no different for advanced work technologies. Overall, people are quite concerned about machines potentially taking jobs. For example, Google-searching "will robots…" yields autofill responses of "take my job," "take over the world," "replace humans," and other concerning outcomes. Further, European residents "overwhelmingly fear job displacement from robots" (Taipale & Fortunati, 2018, p. 201) as do people in nearly every country surveyed (Wike & Stokes, 2018). Anxieties about automation also prompt an introspective bias, whereby people broadly believe automation, robots, or computers will replace human work, yet feel their own jobs are less susceptible (Geiger, 2019).

Views about robots, artificial intelligence (AI), and automation are often related to science-fiction portrayals of robots and artificial intelligence (see Horstmann & Krämer, 2019). For instance, Liang and Lee (2017) found that fear of autonomous robots and artificial intelligence was related to exposure to media and negatively associated with education and income. Using a continent-wide sample of Europeans, Taipale and Fortunati (2018) found that those who were more educated with higher incomes were more likely to see their jobs as more secure in light of increased automation (see also, Turja & Oksanen, 2019). Moreover, income and education are positively related to concerns about automation in representative U.S. samples (Smith, 2016).

## Automation in Pharmacies

Pharmacies, especially large retail chains, are blending "artificial intelligence, machine learning, and advanced software engineering to enhance operational resilience and boost productivity" (BusinessWire, 2020, para. 2). Automation is incorporated in pharmacies in many ways including "record keeping, item selection, labeling, and dose packing" (Spinks et al., 2017, p. 394). Both hospital and retail pharmacies may reconfigure the behind-the-counter workspace for machines (Barrett et al., 2012; Chapuis et al., 2010). But machines are also often placed at centralized distant sites to pre-fill bottles for retail locations (Spinks et al., 2017).

Today, automation in pharmacies is valued at $5.1bn USD and is expected to reach $7.8bn USD by 2024 (Elder, 2019); this increase is largely due to reduced errors and faster fill times. Human error can create a tremendous cost for pharmacies, including unnecessary illness and even loss of life. Human errors occur at a rate of approximately 5 errors per 100,000 orders (Gorbach et al., 2015). But automated prescription filling has error rates near zero (Fanning et al., 2016; cf., Chapuis et al., 2010) and significantly decreases prescription filling time (Walsh et al., 2011).

This bottom-line decision-making focused on operations and productivity likely shapes the organizational reality faced by pharmacy workers, especially those whose roles require less skill (i.e., technicians; Wheeler et al., 2019). Pharmacists and pharmacy technicians have differentiated skill sets, education, and incomes. Thus, pharmacy workers offer a salient sample with differing socioeconomic positions to explore feelings of anxiety tied to automation.

While HMC has been studied during automation implementation in pharmacies (see Barrett et al., 2012), we focus on mental models about automation because mental models guide and are guided by communication (Fairhurst, 2010). Thus, we explore how pharmacy workers think about automation in pharmacy labor. Specifically, we hypothesize differences in automation anxiety and overpayment anxiety will influence perceptions that automation will increase in the future, that pharmacy work will change in the future, and that automation is helpful in pharmacy work.

### Roles in the Pharmacy

Following the SBTC hypothesis, we analyze two divergent socioeconomic groups: pharmacy technicians and pharmacists. Each role contributes to pharmacy work in different ways. *Pharmacy technicians* operate "under the supervision of the licensed pharmacist, [and] assists in pharmacy activities that do not require the professional judgment of a pharmacist" (Albanese et al., 2010, p. e55). The technician's role is juxtaposed alongside robotic interventions: "Along with robotic dispensing technology, [technician] support enables pharmacists to play a more proactive and expanded role in patient care" (p. e36). Pharmacy technicians are typically hourly workers, with less formal training, who assist pharmacists.

Conversely, the role of *pharmacists* is typically broader with a focus on judgment, decision-making, consultation, and supervising. Pharmacists are expected to manage their units, supervise personnel, administer medications, ensure the pharmacotherapy matches the patients' needs, and that the patient understands treatment (Albanese et al., 2010). In addition to formal education, pharmacists are also responsible for maintaining regulated competencies to meet patient needs. In all, though both roles are well-defined, pharmacists are expected to maintain a higher level of skill relative to pharmacy technicians. This hierarchical, educational, monetary, and expertise-driven difference positions pharmacy technicians as lower-skilled workers and pharmacists as skilled experts. If the SBTC hypothesis holds, technology ought to affect pharmacy technicians' mental models more than pharmacists'. Given these discrepancies we propose that in mental models regarding the future of automation and pharmacy work:

> **H1:** Pharmacy technicians will report higher automation anxiety than pharmacists.

> **H2:** Automation anxiety will be positively related to perceptions that (a) the pharmacy job will change across time, (b) automation will increase across time, and (c) automation is helpful to pharmacy work.

Some estimates suggest the probability of pharmacists being replaced by automation is extremely low (.00003) while the probability for replacing pharmacy technicians is quite high (0.92; Frey & Osborne, 2017). Ironically, the O*Net database maintained by the U.S. Department of Labor, which was used as the source for job descriptions in Frey and Osborne's study, classifies pharmacy technicians as a "bright outlook" occupation (onetonline.org, 2020). Frey and Osborne's (2017) study supports the SBTC hypothesis, but it remains unknown how people who work in the same environment, yet have differing skill, pay, and expertise might view automation and the future of their industry. Given the difference in roles taken on by pharmacists and pharmacy technicians, and the predictions by SBTC hypothesis, we ask how these two groups might differ regarding perceptions of future automation in the industry, the benefits of automation, and perceptions that pharmacy jobs will change over time:

> **RQ1:** How do pharmacists and pharmacy technicians differ in terms of (a) perceptions that automation will increase in the coming years, (b) perceived helpfulness of automation, and (c) perceptions that their job will change in the coming years?

Driven by capitalistic market pressures, many industries seek to reduce costs. Technological advancements, like automation, have often been a strategy to deskill work. While this is typically seen as a direct threat to low-income workers, Littler and Innes (2003) found that it also threatens knowledge workers, those with high levels of expertise. In fact, downsizing as a cost-saving effort weakens the job stability of those who are highly compensated and highly educated. This phenomenon has been called deknowledging the workforce and leads to a hollowing out of organizations by thinning out those who are well-paid at the top of the organizational hierarchy, like pharmacists (Frank et al., 2019; Littler & Innes, 2003).

Pharmacists likely view automation as a boon to their work, given the guidance that pharmacists ought to leverage pharmacy technicians and automation processes to boost productivity. However, there has been speculation that dramatic increases in automation processes could threaten pharmacists' jobs as well (Spinks et al., 2017). Spinks et al. point out that automation is typically implemented in a decentralized hub-and-spoke system which physically displaces the jobs remaining. Given the pay discrepancies which pervade the pharmacy environment, it is logical pharmacists might worry about their high pay as a potential liability in light of sophisticated automation. Specifically, while the median wage for pharmacists is greater than $100,000 per year (Polgreen et al., 2011), the median wage for a pharmacy technician is $15 per hour ($31,200 per year for a full-time employee; Wheeler et al., 2019). Further, technicians often work in part-time roles and largely report dissatisfaction with their pay (Desselle & Holmes, 2017). This substantial discrepancy likely affects how both parties view their work and perceive the way automation will affect their career paths. While technicians might perceive pay insufficiency, it is more likely that pharmacists view their salary as a liability. We hypothesize:

> **H3:** Overpayment anxiety will be positively related to perceptions that (a) the pharmacy job will change across time, (b) automation will increase across time, and (c) automation is helpful to pharmacy work.

> **H4:** Pharmacists will experience greater overpayment anxiety than pharmacy technicians.

# Method

## Sample

This research was done in partnership with a state-level pharmacists association. To begin recruitment, the state-level pharmacy association shared the study on its member listserv and permitted us to recruit onsite at their annual conference. We also shared the survey on the Reddit group r/talesfromthepharmacy, visited local pharmacies with flyers, and recruited online with permission on several Facebook and LinkedIn groups dedicated to pharmacy practice. In total we report findings from 131 pharmacists and 109 pharmacy technicians. Just under half of the sample was from Kansas ($n$ = 117, 47.4%) with the remainder from 35 different states.

As expected, pharmacists ($M$ = 11.52, $SD$ = 2.12) and pharmacy technicians ($M$ = 4.95, $SD$ = 2.75) reported very different household incomes (range: 1 = under $10,000 to 13 = $150,000 or more), $t_{corrected}(224) = 2.17$, $p < .001$, $d_{Cohens} = 2.68$. Pharmacists ($M$ = 5.43, $SD$ = 0.90) had higher education levels than pharmacy technicians ($M$ = 3.26, $SD$ = 1.00; 1 = less than H.S. degree, 6 = doctoral degree), $t_{corrected}(224) = 17.73$, $p < .001$, $d_{Cohens} = 2.28$. Three quarters of participants were female ($n$ = 184, 76.7%) and most identified as White ($n$ = 229, 95.4%). Participants' ages ranged from 20 to 80 years ($M$ = 38.15, $SD$ = 12.07).

## Measures

Measures were based on pre-existing items when possible; we began by adding additional items to questions posed by Pew Research (Smith, 2016). To reduce participant burden, and because participants were uncompensated, we used concise measures when possible. Though our measures represent incomplete or limited mental model content, it is reasonable that these measures likely reflect more broadly held assumptions or expectations (Edwards et al., 2019) about machines (e.g., tension, fearfulness, stress, negativity, danger;

Spence et al., 2018). We modified items to fit the pharmacy environment when appropriate. Unless otherwise noted, scales were measured on a Likert-type scale where *1 = strongly disagree* and *7 = strongly agree.* Table 1 presents correlations among demographics and the variables of interest.

**TABLE 1 Correlations, Pooled Means, and SDs, Among Study Variables**

| | M | SD | 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|---|---|---|
| 1. Income | 8.59 | 4.05 | - | | | | | | | |
| 2. Education | 4.38 | 1.41 | 0.66*** | - | | | | | | |
| 3. Age | 38.15 | 12.07 | 0.32*** | -0.13* | - | | | | | |
| 4. Automation Anxiety | 3.28 | 1.49 | -0.19** | -0.17** | -0.03 | - | | | | |
| 5. Overpayment Anxiety | 4.20 | 1.96 | 0.08 | 0.10 | 0.09 | 0.42*** | - | | | |
| 6. Increased Automation | 3.67 | 1.72 | 0.05 | -0.02 | 0.11 | 0.36*** | 0.13* | - | | |
| 7. Job Change | 5.74 | 1.70 | 0.09 | 0.07 | -0.07 | -0.38*** | -0.28*** | -0.26*** | - | |
| 8. Automation Helpfulness | 4.78 | 1.52 | 0.19** | 0.20** | 0.04 | -0.24*** | -0.21** | 0.07 | 0.08 | - |

*Note:* $N = 240$, two participants did not report their age. Each scale is detailed under Sample and Measures above. $p < .05$ *, $p < .01$**, $p < .001$ ***

## Automation Anxiety

To measure feelings of automation anxiety, we used the stem prompt: "I am concerned about losing my job because" with four items: "My employer might use machines or computer programs to replace human workers," "Automation could lead to a reduction in the number of workers needed to do my job," "The job I do today could be done by machines tomorrow," and "Robotic devices could replace people in my role." The scale was reliable, $\alpha = .86$.

## Overpayment Anxiety

We hypothesized that feelings of overpayment anxiety would operate distinctly from automation anxiety. We used the same stem prompt as in automation anxiety with two items: "My employer might find someone who is willing to do my job for less money," and "Someone who will work for less might be hired into my unit." Though these items highly correlated ($R^2 = 0.80$, $p < .001$), the second item was a Heywood case, the indicator had impossible latent-to-item loadings, in the measurement model ($B > 1.00$ among pharmacists; Kline, 2015). The use of only two indicators for a latent factor increases the likelihood of Heywood cases (Kline, 2015). To remedy this, the first item ($M = 4.23$, $SD = 2.09$) was used as a single-item indicator.

## Job Change

Perceptions that one's job will change were measured with the prompt: "Thinking about your current job and occupation, how likely is it that your job will exist in its current form in" with two time frames: 2-years and 5-years. Responses ranged from *1 = very unlikely* to *7 = very likely*. We reverse coded these two items for analysis, they were highly correlated, $R^2 = 0.90$, $p < .001$.

## Increased Automation

We asked participants: "Overall how likely do you think it is that robotic devices, robots, and computers will do much of the work currently done by humans in" with two time frames: 2-years and 5-years. Responses ranged from *1 = very unlikely* to *7 = very likely*. These two items were highly correlated, $R^2 = .86$, $p < .001$.

## Automation Helpfulness

To capture the perceived benefits of automation, we developed three semantic-differential items which used 7-points capturing the perceived helpfulness of automation. Specifically, we prompted participants with "Increased use of automation in my workplace would:" with anchors of "Make my job harder | Make my job easier," "Decrease my efficiency | Increase my efficiency," and "Harm me | Help me." These items were reliable, $\alpha = .92$.

## Measurement Invariance Testing

Given the goals of group synthesis and comparison between pharmacists and pharmacy technicians we first assessed our models for *measurement invariance*, or psychometric equivalence of constructs across groups (Putnick & Bornstein, 2016). We began by creating an unconstrained configural model in the R package lavaan 0.6-5 (Rosseel, 2012). *Configural invariance* occurs when "the latent construct has similar meaning" across groups (Kühne, 2013, p. 155). Configural invariance is established through a similar pattern of loadings across groups and sufficient model fit. The initial configural model revealed a Heywood-case indicator, the second overpayment anxiety item. As noted above, we dropped this item and retained the first. Dropping this item is logical because perceptions that one might be replaced by another who works for less money differ dramatically across pay ranges. Subsequent tests of measurement invariance do not apply to this single item indicator for overpayment anxiety. The baseline configural model yields separate $\chi^2$ values for each group and a pooled model fit; this configural measurement model fit sufficiently well (Kline, 2015): $\chi^2_{\text{Pharmacists}}(90) = 56.07$, $p < .001$, $\chi^2/df = 0.62$, $\chi^2_{\text{Technicians}}(90) = 98.86$, $p < .001$, $\chi^2/df = 1.10$, RMSEA = 0.08, 90% CI[.06, .10], SRMR = 0.04, CFI = 0.96.

Next, we tested *metric invariance,* the assumption that "each item contributes to the latent construct to a similar degree across groups" (Putnick & Bornstein, 2016, p. 75). To test this assumption, indicator loadings were constrained to be equal across the groups before comparing models which were metric invariant: $\chi^2(7) = 4.66$, $p = 0.70$. Third, we tested *scalar invariance*, the assumption the latent constructs have the same meaning and "there is no systematic response bias across populations" (Kühne, 2013, p. 154). This model constrains both indicator loadings and their intercepts and was scalar invariant: $\chi^2(7) = 12.19$, $p = 0.09$. Thus, we proceeded with analysis. We use group comparisons to test H1, RQ1, and H4, then use SEM for H2 and H3.

# Results

## Comparing Pharmacists and Pharmacy Technicians

We first constructed a MANOVA using SPSS 25 to compare pharmacists and pharmacy technicians. Our hypotheses were that pharmacy technicians would have greater automation anxiety (H1), while pharmacists would experience more overpayment anxiety (H4). Further RQ1 asked if these two groups differed in terms of (a) perceptions that automation will increase in the coming years, (b) perceived helpfulness of automation, and (c) perceptions that their job will change in the coming years.

Results of homogeneity of variance as well as normality and multicollinearity assumptions were tested using criteria from Tabachnick and Fidell (2019), no issues were present. The MANOVA was multivariate significant: $F(5, 235) = 7.14$, Wilks' $\lambda = 0.87$, $p < .001$, $\eta^2_{partial} = 0.13$, power = .999. Univariate analysis revealed a significant difference in automation anxiety between pharmacy technicians ($M = 3.54$, $SD = 1.54$) and pharmacists ($M = 3.04$, $SD = 1.43$), $F(1, 239) = 6.82$, $p = .010$, $\eta^2_{partial} = 0.03$. Thus, H1, that pharmacy technicians experience greater automation anxiety, was supported.

However, there were no significant differences between pharmacists and technicians in terms of increased automation in the coming years, $F(1, 239) = 1.13$, $p = 0.289$, $\eta^2_{partial} = 0.01$. Nor was there a difference in terms of perceived job change over time, $F(1, 239) = 0.29$, $p = 0.594$, $\eta^2_{partial} = 0.00$. Perceptions of automation helpfulness did differ between pharmacists ($M = 5.12$, $SD = 1.31$) and technicians ($M = 4.40$, $SD = 1.68$), $F(1, 239) = 13.89$, $p < .001$, $\eta^2_{partial} = 0.06$. Thus, the answer to RQ1a–c is that pharmacists and pharmacy technicians do not differ in terms of perceived future automation or job change over the coming 2 and 5 years. However, pharmacists see automation as more useful than do pharmacy technicians.

The final group comparison was for H4 which contended pharmacists would report higher levels of overpayment anxiety. The univariate tests supported this hypothesis with pharmacists reporting higher overpayment anxiety ($M = 4.48$, $SD = 2.04$) than did technicians ($M = 3.94$, $SD = 2.13$): $F(1, 239) = 3.91$, $p = 0.049$, $\eta^2_{partial} = 0.02$. Thus, H4 was supported.

## Structural Equation Modeling

To test the remaining hypotheses, we conducted an SEM. This model specifies the relationships presented in H2 and H3. We added regression relationships to the measurement model specified above between overpayment and automation anxiety and three outcomes: perceptions that automation will increase, that the job will change in the coming years, and the belief that automation is helpful.

The proposed structural model fit the data sufficiently: $\chi^2_{Pharmacists}(90) = 56.07$, p < .001, $\chi^2/df = 0.62$, $\chi^2_{Technicians}(90) = 98.86$, $p < .001$, $\chi^2/df = 1.10$, RMSEA = .08, 90% CI[.06, .10], SRMR = .04, CFI = 0.96. These values match those presented in the configural measurement model because the relationships among the latent constructs are saturated.

## Pharmacists' Model

Figure 1 shows pharmacists' feelings of anxiety about automation were positively related to perceptions that the job would include increased automation in the future (β = 0.47, *SE* = 0.16, $p < .001$) and positively predicted beliefs that the job would change in the future (β = 0.40, *SE* = 0.13, $p < .001$). Automation anxiety was negatively related to beneficial automation outcomes, (β = –0.24, *SE* = 0.11, $p = .037$). Thus, H2a and b are confirmed among pharmacists. Automation anxieties positively predict perceptions of future automation and job change. Against H2c, pharmacists' automation anxieties were negatively related to perceived benefits of automation.

H3 proposed that overpayment anxieties would predict perceptions that (a) automation will increase in the future, (b) the job will change in the future, and (c) automation is helpful. None of these relationships were significant. Thus, H3 was rejected in the pharmacists' subset.

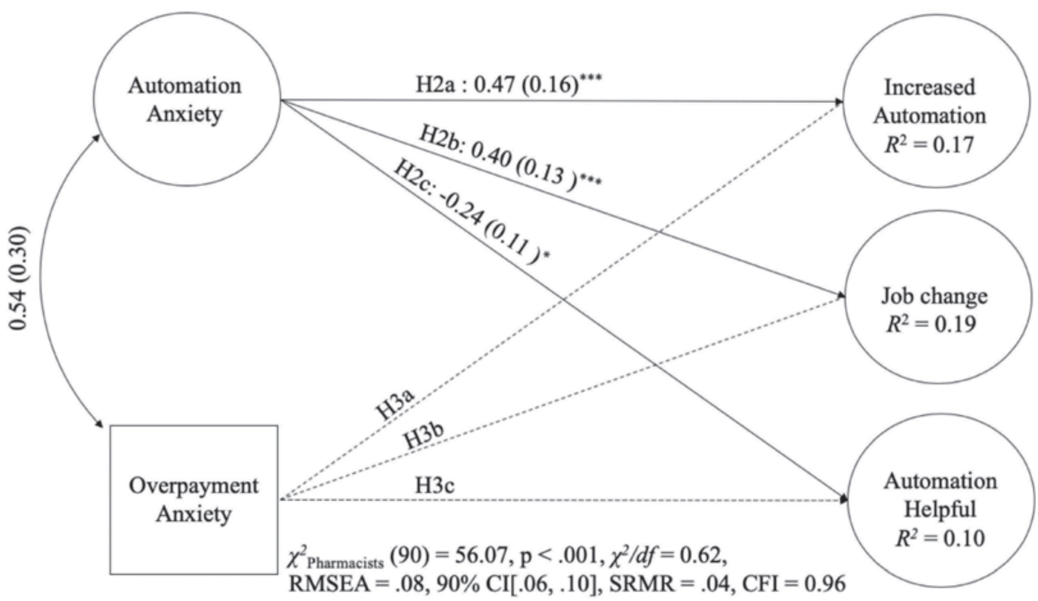



**FIGURE 1 Pharmacists' Model**

*Note: N* = 131 Pharmacists. $p < .05$ *, $p < .001$ ***

## Pharmacy Technicians' Model

Figure 2 shows pharmacy technicians' automation anxiety was positively related to perceptions the job would include increased automation in the future (β = 0.46, *SE* = 0.11, $p < .001$), and positively predicted beliefs that the job would change in the future (β = 0.38, *SE* = 0.12, $p = .001$). However, automation anxiety was unrelated to the perceived helpfulness of automation, ($p = .523$). Thus, H2a and b are confirmed among pharmacy technicians as well. Automation anxieties positively predict perceptions of future automation and job change. Against H2c, technicians' automation anxiety was unrelated to perceived automation benefits.

In the pharmacy technician's model, H3a and b were not supported. However, overpayment anxiety was significantly and negatively related to perceived benefits of automation ($\beta = -0.21$, $SE = 0.08$, $p = .048$). Therefore, pharmacy technicians who felt they were overpaid perceived less benefit from automation in the pharmacy. Thus, H3a–c is rejected in both the pharmacists and pharmacy technician samples. The full technicians' model is presented in Figure 2.

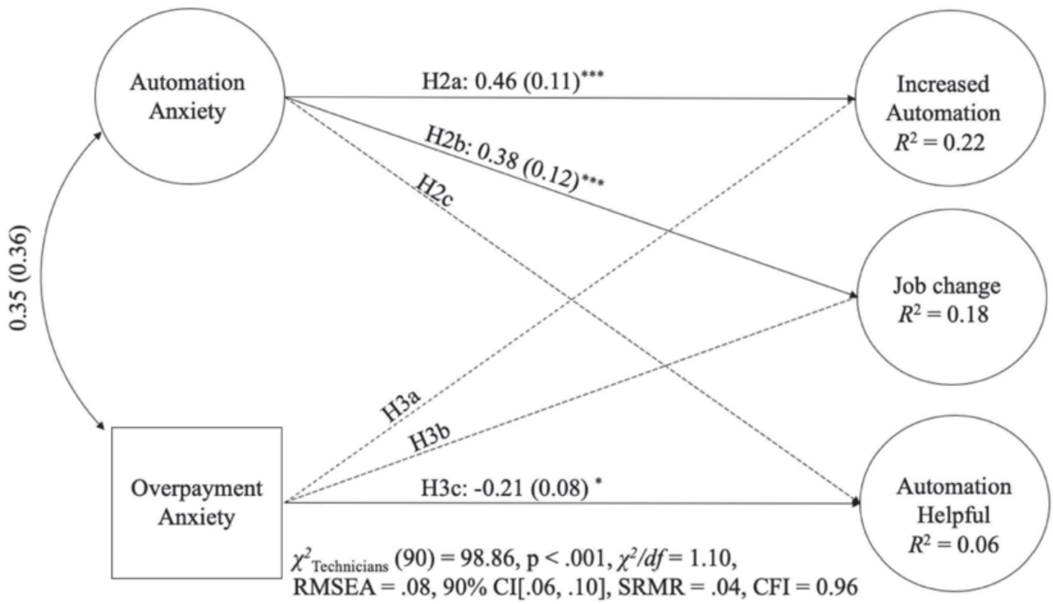



**FIGURE 2 Pharmacy Technicians' Model**

*Note:* $N = 102$ Pharmacy technicians. $p < .05$ *, $p < .001$ ***

## Discussion

This study tests the SBTC hypothesis (Acemoglu & Restrepo, 2019) using a novel sample of pharmacy workers whose education, income, and skills differ by role. Our results juxtapose the SBTC hypothesis against mental models, which drive machine perceptions (Banks, 2020), technology use (Mantovani, 1996), and organizational communication (Fairhurst, 2010). This study adds to previous literature by demonstrating income and education alone are insufficient to determine automation perceptions. Findings reveal pharmacists and pharmacy technicians both experience anxiety related to automation. Despite differing skill, pay, education, and responsibilities, anxiety about automation shapes both parties' view of the future of both automation in pharmacies and future of work in the pharmacy. This is a compelling finding particularly for a sector that is readily adopting automation. Results revealed robust effects for feelings of anxiety about automation with, surprisingly, few effects for overpayment anxiety. These results are intriguing given the differences on key outcomes when comparing pharmacists and pharmacy technicians.

The differences between pharmacists and technicians were significant (accounting for 13% of variance). Pharmacists and technicians seemed to agree that pharmacy automation was likely to increase in the coming years, though this was not significantly different than the scale means for either group. Pharmacists and technicians also agreed that their jobs were likely to change in the coming years, both groups had scale values significantly above the mean ($p < .001$). Pharmacies are a site rich in job change and automation and these findings show that present anxieties about automation translate into beliefs about the future of work and automation.

Pharmacists and technicians did differ in terms of perceived benefits of automation and overpayment anxieties; pharmacists were more likely to feel overpaid and to see automation as beneficial to their work. This is logical given the pay discrepancies and the clear authority role of pharmacists. Interestingly, pharmacists' anxiety about overpayment did not translate into perceptions their work will change or that automation would increase in coming years.

The next take-away from our study is the relatively straightforward mental-model for automation shared by pharmacists and technicians. The models were identical with two exceptions: pharmacy technicians' overpayment anxiety was related to reduced perceptions of automation helpfulness, whereas pharmacists' automation anxieties were related to perceptions of reduced automation helpfulness. Counter to past findings that automation allows pharmacists to reassert "their privileged position . . . while also increasing their institutional legitimacy," (Barrett et al., 2012, p. 1460), these results show both high- and low-skill pharmacy workers share automation anxieties. Feelings of anxiety manifest slightly differently based on position.

We speculate automation anxieties seem to have prompted pharmacists to derogate the benefits of automation (Brehm & Cohen, 1962), whereas overpayment anxieties seem to have the same effect for pharmacy technicians. These findings regarding pharmacists' automation anxiety likely undergird concerns about deknowledging the pharmacy workforce, while pharmacy technicians' overpayment anxiety seems to reflect concerns about deskilling pharmacy work (Littler & Innes, 2003). Thus, it is likely each workgroup derogates automations' helpfulness because of varied job security concerns (reduced knowledge and skill requirements, respectively). Both parties anticipate change over time both in their work in a pharmacy and via increased automation. For both, automation anxiety was related to perceived changes. These results signal that while income and education are forces driving automation anxiety in the general population (Liang & Lee, 2017), in pharmacies, automation anxiety leads to increased beliefs about the changing nature of work and expectations for increased automation. Future, longitudinal, or mixed-methods research could measure and observe perceptions of automation adoption or adaptation over time in order to better understand how mental models change relative to experiences in HMC.

To our knowledge, this is the first test of the SBTC hypothesis based on perceptual data in a recently automated sector. To date, this hypothesis has relied heavily on macro-level models of economic change with little concern about mental models, which shape individuals' everyday communication, interaction, and experiences (Frank et al., 2019). We extend findings from Europe (Taipale & Fortunati, 2018) and the U.S. (Liang & Lee, 2017; Smith, 2016) that income and education buffer fears of automation and include perceptions about the future of work. Our findings reveal this hypothesis has translated across socioeconomic groups into pharmacy workers' beliefs about automation and the changes automation brings to their work.

Further, while other studies have labeled perceptions that automation, computerization, and AI might affect work as a "fear of automation," we choose to label these "feelings of anxiety" to signal the motivational nature of such beliefs. We do not mean to suggest that clinical descriptions of anxiety relate to automation, though they may. Instead, anxiety experienced about major changes to one's work, including the potential for displaced or replaced work, affects the very work practices which combine social and technological counterparts (Barrett et al., 2012). Indeed, feelings of anxiety about new technologies also affects how people adopt, use, and engage with complex machines (Leonardi, 2012). For HMC scholars, these findings reveal how anxieties about machine-communication partners manifest in everyday work practice. It is likely that anxious mental models could foster resistance to HMC, regardless of the benefits technology may offer. Future research would do well to consider the specific manifestations of automation anxiety at work; for example, does anxiety about automation increase/decrease one's effort, satisfaction, or commitment to the workplace? Also, for managers, how might assuring workers' job security amid increased automation affect work outcomes like satisfaction, productivity, or turnover?

For pharmacy workers anxious feelings about automation and the changing nature of work are proximal (i.e., 2 and 5 years) concerns with everyday implications. Indeed, these close temporal estimates about technological and work change were explained quite well by automation anxiety. In addition to the SBTC hypothesis, our study also tests and affirms the proposition that higher-skilled workers also feel vulnerable to technological change (Frank et al., 2019; Littler & Innes, 2003). Future research could benefit from further consideration of how exposure/adoption in specific industries affects feelings of anxiety about automation, especially as interactions with complex technological co-workers increase (Edwards et al., 2019). In retail environments like grocery store checkouts, medical contexts where algorithms aid diagnoses, and practically every other work domain machine counterparts will increase in use (Frey & Osborne, 2017).

The derogation of the helpfulness of automation by pharmacists and technicians was unexpected. However, source and message derogation are logical responses to incompatible messages (Brehm & Cohen, 1962). For these workers, anxieties about automation and overpayment prompted a derogation of the technology that could threaten livelihoods. In line with past research we expect that source and technology derogation affects communication with colleagues and subsequent technological (non)adoption (Leonardi, 2012). To understand how these beliefs form through interaction, future studies might consider networks of HMC interaction.

Given these findings, perceived helpfulness may be tied to outside anxieties and reflect the relational power dynamics between organizational decision-makers and front-line workers. As shown in Table 1, perceptions of helpfulness are negatively correlated with both automation anxiety and overpayment anxiety. Yet, perceptions of helpfulness were not significantly related to perceptions that the job will change and automation will increase. As industries increasingly adopt automation, it is crucial for decision makers to be aware of potential anxieties that can reduce the perceived helpfulness of technology. For managers hoping to implement automation, this finding signals a need to assess perceptions that the technologies might replace workers as part of the persuasive campaign to increase adoption of automated technologies. Certainly, this validates the argument by HMC scholars that perceptions of technological partners have meaningful consequences for our subsequent communication with and about such technologies (Fortunati & Edwards, 2020); indeed, this argument seems salient for the impending future of work.

## Conclusion and Limitations

This study extends our understanding of how technological change prompts feelings of anxiety among both high- and low-skilled workers. While our study offers insights to a racially White but relatively diversified socioeconomic groups, in a recently automated sector, the specific context also limits the generalizability of these findings. For these workers feelings of automation anxiety predicted variance in perceptions of the near future and the helpfulness of automation. The measures were limited by relatively few items for each construct. Mental models represent an abstract preconception, certainly these mental models could be tied to a wide variety of antecedents and experiences. Future research will benefit from mapping the constellation of both the content and structure of these mental models, beyond the SBTC hypothesis, to better understand antecedents and outcomes for technological concern (see also Banks, 2020).

In all, this study enhances scholarly understanding of how feelings of anxiety affect cognitive mental models, which inform HMC. The future will inevitably include increased automation; thus, perceptions of that process deserve continued scholarly attention. We argue scholars and practitioners ought to attend to automation and overpayment anxieties as they directly connect to feelings about the future and the mental models which permeate everyday communication and work, especially in complex HMC situations.

## Author Biographies

**Cameron W. Piercy** (PhD) is an assistant professor in the Department of Communication Studies at the University of Kansas and the founding director of the Human-Machine Communication Lab (https://hmc.ku.edu). His research explores how technology is (re) shaping social networks. He is especially interested in how perceptions of technology affect the way people engage with both technologies and other people. His research has been published in *Communication Research, Computers in Human Behavior,* and *New Media & Society,* among others.

https://orcid.org/0000-0003-1431-3086

**Angela N. Gist-Mackey** (PhD) is an assistant professor in the Department of Communication Studies at the University of Kansas with expertise in organizational communication. She is largely an interpretive critical scholar who researches issues of social mobility, power, and organizing. Her program of research frequently explores topics related to social class and various stigmatized social identities. Much of Dr. Gist-Mackey's research is engaged, qualitative scholarship partnering with local organizations to provide recommendations for their programs. Her scholarship has been published in outlets such as *Organization Studies, Communication Monographs, Journal of Applied Communication Research,* and *Management Communication Quarterly,* among others

https://orcid.org/0000-0002-1590-0728

## References

Acemoglu, D., & Restrepo, P. (2019). Automation and new tasks: How technology displaces and reinstates labor. *Journal of Economic Perspectives*, *33*(2), 3–30. https://doi.org/10.1257/jep.33.2.3

Albanese, N. P., Rouse, M. J., & Schlaifer, M. (2010). Scope of contemporary pharmacy practice: Roles, responsibilities, and functions of pharmacists and pharmacy technicians. *Journal of the American Pharmacists Association*, *50*, e35–e69. https://doi.org/10.1331/JAPhA.2010.10510

Autor, D. H. (2015). Why are there still so many jobs? The history and future of workplace automation. *Journal of Economic Perspectives*, *29*(3), 3–30. https://doi.org/10.1257/jep.29.3.3

Banks, J. (2020). Optimus primed: Media cultivation of robot mental models and social judgments. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.00062

Barrett, M., Oborn, E., Orlikowski, W. J., & Yates, J. (2012). Reconfiguring boundary relations: Robotic innovations in pharmacy work. *Organization Science, 23,* 1448–1466. https://doi.org/10.1287/orsc.1100.0639

Baym, N. K. (2015). *Personal connections in the digital age*. John Wiley & Sons.

Berman, E., Bound, J., & Machin, S. (1998). Implications of skill-biased technological change: International evidence. *The Quarterly Journal of Economics*, *113*, 1245–1279. https://doi.org/10.1162/003355398555892

Brehm, J. W., & Cohen, A. R. (1962). *Explorations in cognitive dissonance*. Wiley & Sons.

Brynjolfsson, E., & Mitchell, T. (2017). What can machine learning do? Workforce implications. *Science*, *358*, 1530–1534. https://doi.org/10.1126/science.aap8062

BusinessWire. (2020, February 3). *Walgreens boots alliance advances transformation of its global information technology operating model to accelerate digitalization, drive efficiencies and savings*. https://web.archive.org/web/20200713130332/https://www.businesswire.com/news/home/20200203005432/en/Walgreens-Boots-Alliance-Advances-Transformation-Global-Information

Chapuis, C., Roustit, M., Bal, G., Schwebel, C., Pansu, P., David-Tchouda, S., Foroni, L., Calop, J., Timsit, J.-F., Allenet, B., Bosson, J.-L., & Bedouch, P. (2010). Automated drug dispensing system reduces medication errors in an intensive care setting. *Critical Care Medicine*, *38*, 2275–2281. https://doi.org/10.1097/CCM.0b013e3181f8569b

Colbert, A., Yee, N., & George, G. (2016). The digital workforce and the workplace of the future. *Academy of Management Journal*, *59*, 731–739. https://doi.org/10.5465/amj.2016.4003

Desselle, S. P., & Holmes, E. R. (2017). Results of the 2015 national certified pharmacy technician workforce survey. *American Journal of Health-System Pharmacy*, *74*, 981–991. https://doi.org/10.2146/ajhp160666

Edwards, A., Edwards, C., Westerman, D., & Spence, P. R. (2019). Initial expectations, interactions, and beyond with social robots. *Computers in Human Behavior*, *90*, 308–314. https://doi.org/10.1016/j.chb.2018.08.042

Elder, M. (2019, September). *Pharmacy Automation Technology Market Size and Research Report*. BCC Research. https://web.archive.org/web/20200713135920/https://www.bccresearch.com/market-research/instrumentation-and-sensors/pharmacy-automation-tech-markets-report.html

Fairhurst, G. T. (2010). *The power of framing: Creating the language of leadership*. John Wiley & Sons.

Fanning, L., Jones, N., Grand Dip Clinical Pharmacy, & Manias, E. (2016). Impact of automated dispensing cabinets on medication selection and preparation error rates in an emergency department: A prospective and direct observational before-and-after study. *Journal of Evaluation in Clinical Practice*, *22,* 156–163. https://doi.org/10.1111/jep.12445

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, *1*. https://doi.org/10.30658/hmc.1.1

Frank, M. R., Autor, D., Bessen, J. E., Brynjolfsson, E., Cebrian, M., Deming, D. J., Feldman, M., Groh, M., Lobo, J., Moro, E., Wang, D., Youn, H., & Rahwan, I. (2019). Toward understanding the impact of artificial intelligence on labor. *Proceedings of the National Academy of Sciences*, *116*, 6531–6539. https://doi.org/10.1073/pnas.1900949116

Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, *114*, 254–280. https://doi.org/10.1016/j.techfore.2016.08.019

Gebhart, F. (2019, July 4). The Future of Pharmacy Automation. *Drug Topics*. https://web.archive.org/web/20201010183848/https://www.drugtopics.com/view/future-pharmacy-automation

Geiger, A. W. (2019, April 8). *How Americans see automation and the workplace | Pew Research Center*. Pew Research. https://web.archive.org/web/20190530220726/https://www.pewresearch.org/fact-tank/2019/04/08/how-americans-see-automation-and-the-workplace-in-7-charts

Gorbach, C., Blanton, L., Lukawski, B. A., Varkey, A. C., Pitman, E. P., & Garey, K. W. (2015). Frequency of and risk factors for medication errors by pharmacists during order verification in a tertiary care medical center. *American Journal of Health-System Pharmacy*, *72*, 1471–1474. https://doi.org/10.2146/ajhp140673

Green, M. C., & Clark, J. L. (2015). Real or ersatz? Determinants of benefits and costs of online social interactions. In S. Sundar (ed.), *The Handbook of the Psychology of Communication Technology* (pp. 247–269). Wiley Blackwell. https://doi.org/10.1002/9781118426456.ch11

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00939

Hynniman, C. E., & Lamy, P. P. (1967). Outpatient pharmacy automation. *American Journal of Hospital Pharmacy*, *24*, 18–21. https://doi.org/10.1093/ajhp/24.1.18

Kline, R. B. (2015). *Principles and practice of structural equation modeling*. Guilford Press.

Kühne, R. (2013). Testing measurement invariance in media psychological research. *Journal of Media Psychology*, 25, 153–159. https://doi.org/10.1027/1864-1105/a000096

Leonardi, P. M. (2012). *Car crashes without cars: Lessons about simulation technology and organizational change from automotive design.* MIT Press.

Liang, Y., & Lee, S. A. L. (2017). Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics*, *9*, 379–384. https://doi.org/10.1007/s12369-017-0401-3

Littler, C. R., & Innes, P. (2003). Downsizing and deknowledging the firm. *Work, Employment and Society*, *17*, 73–100. https://doi.org/10.1177/0950017003017001263

Madden, E. E., & Dreyfus, R. H. (1968). Outpatient pharmacy prescription automation. *American Journal of Hospital Pharmacy*, *25*, 20–25. https://doi.org/10.1093/ajhp/25.1.20a

Mantovani, G. (1996). Social context in HCl: A new framework for mental models, cooperation, and communication. *Cognitive Science, 20*, 237–269. https://doi.org/10.1207/s15516709cog2002_3

O*Net. (2020, July 7). Summary report for: 29-2052.00-Pharmacy Technicians. https://web.archive.org/web/20200528144913/https://www.onetonline.org/link/summary/29-2052.00

Polgreen, L. A., Mott, D. A., & Doucette, W. R. (2011). An examination of pharmacists' labor supply and wages. *Research in Social and Administrative Pharmacy*, *7*(4), 406–414. https://doi.org/10.1016/j.sapharm.2010.10.006

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*, 1–36. https://doi.org/10.18637/jss.v048.i02

Smith, A. (2016). *Public Predictions for the Future of Workforce Automation* Pew Research. https://web.archive.org/web/20210110192739/https://www.pewresearch.org/internet/2016/03/10/public-predictions-for-the-future-of-workforce-automation

Spence, P. R., Westerman, D., & Lin, X. (2018). A robot will take your job. How does that make you feel? Examining perceptions of robots in the workplace. In A. L. Guzman (Ed.), *Human-Machine Communication: Rethinking communication, technology, and ourselves* (pp. 185–200). Peter Lang.

Spinks, J., Jackson, J., Kirkpatrick, C. M., & Wheeler, A. J. (2017). Disruptive innovation in community pharmacy – Impact of automation on the pharmacist workforce. *Research in Social and Administrative Pharmacy*, *13*, 394–397. https://doi.org/10.1016/j.sapharm.2016.04.009

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Taipale, S. de Luca, F., Sarrica, M., & Fortunati, L. (2015). Robot shift from industrial production to social reproduction. In J. Vincent, S. Taipale, B. Sapio, G. Lugano, & L. Fortunati (Eds.) *Social Robots from a Human Perspective* (pp. 11–24). Springer.

Taipale, S., & Fortunati, L. (2018). Communicating with machines: Robots as the next new media. In A. L. Guzman (Ed.), *Human-Machine Communication: Rethinking communication, technology, and ourselves* (pp. 201–219). Peter Lang.

Turja, T., & Oksanen, A. (2019). Robot acceptance at work: A multilevel analysis based on 27 EU countries. *International Journal of Social Robotics*, *11*, 679–689. https://doi.org/10.1007/s12369-019-00526-x

Walsh, K. E., Chui, M. A., Kieser, M. A., Williams, S. M., Sutter, S. L., & Sutter, J. G. (2011). Exploring the impact of an automated prescription-filling device on community pharmacy technician workflow. *Journal of the American Pharmacists Association*, *51*(5), 613–618. https://doi.org/10.1331/JAPhA.2011.09166

West, D. M. (2018). *The future of work: Robots, AI, and automation*. Brookings Institution Press.

Wheeler, J. S., Renfro, C. P., Wang, J., Qiao, Y., & Hohmeier, K. C. (2019). Assessing pharmacy technician certification: A national survey comparing certified and noncertified pharmacy technicians. *Journal of the American Pharmacists Association*, *59*(3), 369–374.e2. https://doi.org/10.1016/j.japh.2018.12.021

Wike, R., & Stokes, B. (2018). *In advanced and emerging economies alike, worries about job automation.* Pew Research. https://web.archive.org/web/20200713140521/https://www.pewresearch.org/global/wp-content/uploads/sites/2/2018/09/Pew-Research-Center_In-Advanced-and-Emerging-Economies-Alike-Worries-about-Job-Automation_2018-09-13.pdf

Human-Machine Communication
Volume 2, 2021
https://doi.org/10.30658/hmc.2.11

# Out With the Humans, in With the Machines?: Investigating the Behavioral and Psychological Effects of Replacing Human Advisors With a Machine

**Andrew Prahl**[1] and **Lyn M. Van Swol**[2]

1 Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore
2 Department of Communication Arts, University of Wisconsin-Madison, Madison, WI, USA

## Abstract

This study investigates the effects of task demonstrability and replacing a human advisor with a machine advisor. Outcome measures include advice-utilization (trust), the perception of advisors, and decision-maker emotions. Participants were randomly assigned to make a series of forecasts dealing with either humanitarian planning (low demonstrability) or management (high demonstrability). Participants received advice from either a machine advisor only, a human advisor only, or their advisor was replaced with the other type of advisor (human/machine) midway through the experiment. Decision-makers rated human advisors as more expert, more useful, and more similar. Perception effects were strongest when a human advisor was replaced by a machine. Decision-makers also experienced more negative emotions, lower reciprocity, and faulted their advisor more for mistakes when a human was replaced by a machine.

**Keywords:** human-machine communication, interpersonal communication, advice, task demonstrability, emotion

## Introduction

On August 13, 2014, a video titled *Humans Need Not Apply* was uploaded to YouTube and exploded in popularity, gathering over 1 million views within 3 days; it currently stands at

**CONTACT** Andrew Prahl • Wee Kim Wee School of Communication & Information • Nanyang Technological University • 31 Nanyang Link • Singapore 637718 • andrew.prahl@ntu.edu.sg

ISSN 2638-602X (print)/ISSN 2638-6038 (online)
www.hmcjournal.com

over 10 million views (CGPGrey, 2014; Pagano, 2014). Detailing a future in which human labor is irrecoverably taken over by automation, the 15-minute video on the evolution of labor—and why the robot revolution is different—was described as "terrifying" and foretelling "an economic horror movie" by some commentators (Roggeveen, 2014, p. 1). Since then, public interest in the continued automation of human labor has only increased; it is becoming difficult to read the business section of a popular newspaper and *not* encounter an article discussing the future of work and automation. Emerging alongside the increased discussion of machines versus humans, human-machine communication is a quickly growing subfield of communication which studies machines as interlocutors rather than simply as communication tools (Banks & de Graaf, 2020; Fortunati & Edwards, 2020). This study draws upon both interpersonal and human-machine trust literature to investigate a common real-world use for new automation: acting as an advisor to a human decision-maker. Industries as diverse as health care (Langlotz et al., 2019), finance (Lourenço et al., 2020), supply-chain management (Fildes & Goodwin, 2020), and agriculture (Zhai et al., 2020) are increasingly turning toward machines as advisors. In financial advising, for example, robo-advisors currently manage an estimated $1 trillion in assets, a number that is expected to increase to over $15 trillion by 2025 (Abraham et al., 2019; Deloitte, 2016).

The interpersonal process of trust has attracted considerable interest from scholars in a wide variety of fields (for review, see Bonaccio & Dalal, 2006). More recently, the study of human-machine trust has also increased (Hoff & Bashir, 2015; Lutz & Tamò-Larrieux, 2020). There is also scholarship emerging that attempts to explain the differences in how human and machine communicators are conceptualized (Guzman, 2020). However, little research has experimentally compared human-machine trust directly to interpersonal trust, especially in situations where machines replace humans or for decisions with more subjective and less demonstrable consequences. Given that the thought of machines replacing humans concerns many people, it is surprising how little research exists on the psychological state of people who witness machines replacing humans. The purpose of this study is to investigate (1) the effects of task demonstrability on trust in humans and machines, (2) how perceptions of advisors are affected by task type, advisor type, and advisor replacement. We begin by conceptualizing differences in task demonstrability and advisor expertise.

## Task Demonstrability and Advisor Expertise

A continuum of decision-making task demonstrability anchored by intellective (high demonstrability) and judgmental tasks (low demonstrability) was explicated by Laughlin and Ellis (1986). This theoretical distinction has become important in advice research (Bonaccio & Dalal, 2006). Demonstrable tasks are distinguished by having an answer that all parties can understand. For example, an algebra problem has a correct answer and any advice provided to a decision-maker suggesting a correct answer is demonstrably correct or incorrect. On the other hand, low demonstrability tasks involve uncertain future states or subjective consequences, and the decision is seen more as a value judgment than a correct answer. Several interpersonal advice studies have varied decision-making tasks on the demonstrability continuum (Tzioti et al., 2014; Van Swol, 2011).

One of the strongest and most robust effects in advice research is that perceptions of advisor expertise directly affect trust in the advisor (Sniezek & Van Swol, 2001). Given that task demonstrability and advisor expertise are key constructs in interpersonal advice research, human-machine communication provides an ideal context to study how the two factors may interact because the expertise of an advisor is tied to the domain and decision-making context. This suggests that perceptions of expertise for low or high demonstrability of tasks may differ for humans and machines. In the research presented here, we manipulate both advisor type (human/machine) and demonstrability of the decision-making task. Decades of research on technology acceptance, human-automation (human-machine) trust, and the growing field of machine ethics provides some theoretical insight as to how perceptions of machine attributes differ from humans.

## Perceptions of Machines and Perceptions of Humans

The comparison of interpersonal and human-machine trust in tasks of varying demonstrability, especially those involving moral decisions, introduces an interesting question about the match between advisor characteristics and the context of a decision. The majority of human-machine advice research has used highly demonstrable tasks as experimental stimuli, such as what the next number is in a mathematical sequence (de Visser et al., 2016) or a yes/no question regarding the presence of military equipment in an aerial surveillance photo (Rice & Geels, 2010). There exists little research to guide our assumptions about trust of machine advisors on less demonstrable tasks. However, machines may need to make such decisions in the future, such as a self-driving car that must decide to protect vehicle occupants at the cost of endangering pedestrians (Awad et al., 2018). Although not focused on specific decisions, some scholars have investigated perceptions of various forms of machines (e.g., robots) for varying roles in society (Katz & Halpern, 2014; Takayama et al., 2008). Machines are often perceived as being more suitable for roles that do not require emotion or sensitive communication, and more suitable for roles that require memorization and unselfish service-orientation (Takayama et al., 2008).

Several perspectives in human-machine communication literature highlight the importance of the different expectations between humans and machines (Gambino et al., 2020; Guzman, 2020; Madhavan & Wiegmann, 2007). For example, according to the perfect automation schema model (Madhavan & Wiegmann, 2007), automation is expected to be high performing but invariant, whereas humans (if not as high performing) are adaptable and are expected to learn from their mistakes. Additionally, mistakes are less expected from machines in general because decision-makers do not see machines as susceptible to biases and emotions that plague human judgment (Merritt et al., 2015). On the other hand, people know that other people are not perfect. Several authors in human-robot trust have investigated questions of emotion, but there is only limited evidence that humans consider machines to possess emotion (Guzman, 2020; Kahn et al., 2012).

If human and machine advisors are perceived as having differing fundamental attributes, it will affect the perception of either advisor's capabilities. It follows that when an advisor is not assessed to have expertise (i.e., capabilities required for a certain task), the assessments of the advisor that are dependent on the task will be affected as well. Because less demonstrable tasks require value judgments that, in turn, are tied to emotion and subjective evaluation (Horberg et al., 2011), we believe that the underlying assumption that machines lack emotion will lead to lower assessment of advisor expertise in less demonstrable tasks.

**Hypothesis 1a:** Machine advisors will be perceived as having less expertise in less demonstrable decision tasks than more demonstrable tasks.

Perceptions of advisor expertise also affect perceptions of the advice itself. If advice comes from non-expert sources, it is perceived as less useful and less appropriate. This effect is both logical and established in past advice research (Bonaccio & Dalal, 2006).

**Hypothesis 1b/c:** Advice from machine advisors will be perceived as being less (1b) useful (1c) appropriate in less demonstrable than more demonstrable decision tasks.

## Task, Advisor, Trust, and Perception

In the experiment below, we attempt to control factors like task that moderate the relationship between perceptions of the advice and advisor expertise, therefore we expect to see the effects of hypotheses 1a–1c reflected in our behavioral trust measure (advice-utilization) as well, especially as advice utilization and perception of advice are measured closely in time and previous research has found a strong relationship between perception of advice and use of advice (Bonaccio & Van Swol, 2014).

**Hypothesis 2:** Machine advice will be utilized less than human advice, in general, when the decision is less demonstrable.

Our third hypothesis is reasoned from hypotheses 1 and 2. Because different tasks may result in different advisor characteristics becoming more salient, we predict that recipient perceptions of thought process and value similarity with human advisors will be higher in less demonstrable decision tasks because the task is thought to be better suited to a decision-maker that possesses emotion.

**Hypothesis 3a/b:** Human/machine advisors will be perceived as having more (3a) thought process and (3b) value similarity in less demonstrable decision tasks.

## Advisor Replacement

In the research study below, we manipulate advisor type, task type, and advisor replacement. There is essentially no interpersonal or machine trust research that has been conducted specifically to test the effects of advisor replacement, but because perception of humans and machine relies on different underlying assumptions, perceptual effects related to the comparison of two stimuli may be applicable to guide our expectations. When one advisor is replaced by another type, this may elicit a comparison of the two advisors that makes the perceived attributes of both more salient. Such an effect would fit into existing literature on contrast effects in communication and impression formation research (Palmer & Gore, 2014).

Contrast effects describe the process by which exposure to one target of evaluation can change the evaluation of targets presented subsequently. For example, unattractive faces are rated as more unattractive if the evaluator is shown an attractive face before the unattractive one (Wedell et al., 1987). We are not aware of any literature to suggest that contrast effects will not extend to the evaluation of one advisor after replacing another. If a contrast effect is found, we expect it to result in our hypothesized effects of advisor type becoming stronger. For example, if a decision-maker is presented with a new machine advisor after gaining experience with a human advisor, it may result in an even stronger perception of invariance and exaggerate expectations of high performance. To be clear, we only manipulate advisor replacement, we do not replace one task with another; our hypothesis below therefore only covers effects driven by advisor perception. For brevity, we summarize these effects in the below hypotheses:

> **Hypothesis 4a/b:** Machine/human advisors will be evaluated as less expert, useful, appropriate, and similar in less/more demonstrable decision tasks when they replace human/machine advisors than when replacing another machine/human advisor.

> **Hypothesis 4c/d:** In more/less demonstrable tasks, when a machine/human advisor replaces a human/machine advisor, the machine/human advice will be utilized more than when a machine/human advisor replaces another machine/human advisor.

## Decision-Maker Emotions

Our earlier discussion of emotions primarily discussed a decision-maker's perception that an advisor possesses emotions or at least the capability to understand emotions. But perceptual processes themselves are affected by emotions, and interpersonal advice research has shown decision-maker emotions to have substantial effects on trust (MacGeorge et al., 2013). In interpersonal advice research using demonstrable tasks, researchers have manipulated decision-maker emotions, finding that the induction of other-directed negative emotions (i.e., anger, frustration) resulted in less advice utilization, while other-directed positive emotions (i.e., happiness, gratitude) resulted in more utilization (Gino & Schweitzer, 2008). Such effects were also found in research using less demonstrable tasks (de Hooge et al., 2014).

Research on the effects of decision-maker emotions and trust in machines is less conclusive about the effects of emotions on trust. This is largely because advisor anthropomorphism can have strong effects on emotion (de Visser et al., 2016; Waytz et al., 2014), and there are large differences in anthropomorphism for machines (i.e., a social robot versus a calculator). Because we do not manipulate decision-maker emotion in our study, we are only able to predict potential effects that result from our manipulations advisor type and task, but our study design is ideally suited to investigate the emotions that may be produced by interacting with human versus machine advisors and the effects of replacing one advisor with another. A human being replaced by a machine, for example, could produce a negative emotional reaction due to the belief that the machine is not suited for the decision task or vice versa. Thus, we incorporate measures of positive emotions (e.g., happiness) and negative emotions (e.g., anger).

In addition to emotions, we also measure two processes related to trust: reciprocity and fault. Reciprocity—the belief of owing something to one's advisor—is interesting because trust is often conceptualized as a reciprocal process (Mayer et al., 1995). We also examine attributions of fault for mistakes because if humans are expected to be fallible and imperfect (Madhavan & Wiegmann, 2007), it may result in decision-makers generally finding less fault in human advisors' mistakes. We are also interested in fault because it is possible that decision-makers will fault machines to a greater degree than human advisors because fault may be related to *blame*. Our low demonstrability decision task in this experiment has consequences that result in the loss of human life, and though the measurement of what is perceived as "moral" is complicated, it is not unreasonable to assume that decision-makers could sense moral implications. Some research and emerging machine ethics research suggests many humans have a discomfort with placing blame on machines for making decisions with moral implications because many people do not perceive machines to possess moral accountability (Kahn et al., 2012), our experimental manipulations offer a unique opportunity to investigate this question.

> **RQ1/2/3:** Are decision-maker (1) emotions, (2) reciprocity, and (3) attributions of fault affected by task and advisor type?

## Method

### Participants

Participants were recruited through Amazon's Mechanical Turk (MTurk) service and were required to be U.S. citizens over 18. Typical MTurk samples have limitations; for example, they tend to be younger and more likely to vote Democratic (Levay et al., 2016), and the use of scripts or bots is possible. To minimize potential problems, we first specified that subjects were "Master" workers (have a history of providing high quality work), and we used MTurk worker qualifications (i.e., age, gender, geographic location) to ensure a sample similar to the U.S. general population. Finally, our screening questions on the survey itself were set with quotas of demographics such as age and gender as a second layer of verification. Throughout the survey, attention and bot check questions were presented at random intervals; any subject failing two or more check questions was eliminated. Power analyses were conducted based off of past research (see appendix) and given the very slight manipulations present in our research, we recruited a large enough sample to detect effects. A total of 689 participants completed the study. In the high demonstrability task: $n = 321$, there were $n = 80$ participants in the machine advisor replaced by machine (MrM), $n = 77$ in human advisor replaced by human (HrH), $n = 84$ in machine advisor replaced by human advisor (MrH), $n = 80$ in human advisor replaced by machine advisor (HrM). Low demonstrability task: $n = 368$, $n = 82$ (MrM), $n = 74$ (HrH), $n = 104$ (MrH), $n = 108$ (HrM).

### Task

A forecasting task was chosen to maximize relatability to previous research comparing human and machine advice (e.g., Fildes et al., 2006; Önkal et al., 2009), because it allowed for the clean manipulations of task demonstrability, and because forecasting is a task for which machines are being increasingly used in the real world, for example, in supply chain forecasting (Fildes & Goodwin, 2020). Participants completed 20 forecasting tasks; all 20 graphs/forecast scenarios were randomly assigned within each task condition. In the high task demonstrability condition, the forecasting scenarios related to hospital operating room management (screenshots in appendix). In the low task demonstrability condition, the same graphs were displayed, but the scenarios dealt with humanitarian relief.[1]

## Procedure

The manipulation of advisor type was simple and similar to past studies (Önkal et al., 2009; Prahl & Van Swol, 2017). Participants were told at the opening that the advice would come either from an algorithmic software program (OptiLytics), or an experienced surgeon at the hospital in the high demonstrability condition (in the low demonstrability condition, a humanitarian relief professional). The advisors were introduced to participants with a short photograph describing their/its role in the organization (see appendix for descriptions). Midway (after trial 10), the advisor was replaced with either the same type (human/machine) or different type of advisor to create the HrH, HrM, MrM, and MrH conditions.[2]

Compensation for participants was set at 25% above the federal (USA) minimum wage rate assuming a 45-minute completion time. Participants were presented graphical representations of past data, similar to a stock price chart, and then asked to make an initial forecast of where the value would be in the future. After making an "initial" forecast, the advice from a human (or algorithmic) advisor was presented; participants could make a "revised" forecast on the screen and submit it as their final forecast.[3]

At the end of each trial, a performance feedback screen was shown that displayed the participant's final forecast, the advisor's forecast, and the actual correct answer. Additionally, percentage errors were calculated for each forecast (allowing them to compare their own performance versus the advisor's performance). The participant was also shown their average percentage error across all trials so they could see if their performance on the individual trial was better or worse than previously. Finally, in the high demonstrability condition, the participant's percentage error was multiplied by 1000 and presented as (for a 1% error): "This forecasting error is estimated to have cost the hospital $1000." In the low demonstrability condition, the percentage error was multiplied by 100 and displayed as: "This forecasting error is estimated to have resulted in 100 deaths." This was to reinforce that the decisions had either financial consequences or consequences resulting in adversity to humans. The first 10 trials were performed to set the stage for the second group of 10 trials where our research interest in advisor replacement lies.

1. We picked these domains partially to control for participants having personal intuition for what the outcomes would be. Recipients often use advice less if they believe themselves to possess unique domain expertise (Lawrence et al., 2006); our task minimizes this risk.

2. When advisors were replaced midway through the survey, the introduction text was preceded by "Due to time constraints, we were not able to get [advisor]'s advice for every forecast. Therefore, you will have a new advisor to help you on the remaining tasks. Your new advisor is . . . "

3. Javascript coding was written into the survey to control the accuracy of the participant forecast and advice forecast, which was varied slightly between trials. This resulted in the relative performance of the participant and the advisor always being the same across all participants to control for related confounds.

### Measures

A measure used in past research to assess behavioral trust, the "SHIFT" variable, was used to measure advice-utilization. This measure not only provides commonality with forecasting research and algorithmic advice research (e.g., Önkal et al., 2009), but also commonality with interpersonal advice studies which have used the equivalent "Weight of Advice" measure to assess trust as a behavioral measure (for review, see Bonaccio & Dalal, 2006). The SHIFT formula is:

> (Judge revised forecast – Judge initial forecast) / (Advisor forecast – Judge initial forecast)

A questionnaire was administered to assess perceptions of the advisor (e.g., expertise), advice (e.g., appropriateness), and decision-maker emotions, reciprocity, and fault after the first set of 10 trials and after the final 10 trials with the replacement advisor.[4] The survey, consisting of sematic differential and Likert style survey questions, was constructed with survey items from previous advice literature (MacGeorge et al., 2013), details of survey measures and reliability can be found in the appendix.

## Results

### Manipulation Checks

To confirm participants perceived the tasks differently, we conducted independent sample *t* tests on our manipulation check questions between tasks. Means and standard deviations are displayed in Table 1. The task manipulation was successful.

### Hypotheses

Hypothesis 1 stated that machine advice would be perceived as (1a) less expert, (1b) less useful, and (1c) less appropriate than human advice in the humanitarian than the management task. We conducted two-way ANOVAs for perceptions of advisor expertise, appropriateness of advice, and usefulness of advice in the first block (pre-replacement) with task type and advisor type as independent variables. There was no significant interaction between task type and advisor type for expertise of advisor $F(1,669) = 0.062$, $p = 0.803$, $d = 0.062$, appropriateness of advice, $F(1,658) = 0.025$, $p = 0.873$, $d = 0.058$, or usefulness of advice, $F(1,658) = 0.003$, $p = 0.998$, $d = 0.001$. Follow-up univariate tests indicated a main effect of advisor only: human advice was always perceived as being more expert $F(1,669) = 20.681$, $p < 0.001$, $d = 0.356$, more appropriate $F(1,658) = 53.803$, $p < 0.001$, $d = 0.570$, and more useful $F(1,658) = 62.350$, $p < 0.001$, $d = 0.616$, in the first block of trials. Because the main effect of advisor type was significant in the survey results from the first block of trials,

4. The questionnaire measures were identical and measured perceptions of advice usefulness and advisor quality on a sematic differential scale. Additionally, Likert survey questions measured emotions when receiving advice, trust of advisor, similarity (value, social norm, and thought process) to advisor, and perceptions of advisor effort.

**TABLE 1 Manipulation Checks Independent *t* Tests**

| Statement | Humanitarian Task | Management Task | Mean Difference | *p* | *d*** |
|---|---|---|---|---|---|
| Task is more about Human Life | *M* = 5.573<br>*SD* = 1.381 | *M* = 2.938<br>*SD* = 1.554 | 2.634 | 0.001* | 1.845 |
| Task is more about Money | *M* = 2.781<br>*SD* = 1.283 | *M* = 4.008<br>*SD* = 2.190 | 1.226 | 0.001* | 1.064 |
| Task has no right answers | *M* = 2.641<br>*SD* = 1.269 | *M* = 2.349<br>*SD* = 1.284 | 0.293 | 0.004 | 0.251 |
| Task requires compassion | *M* = 3.424<br>*SD* = 1.761 | *M* = 3.058<br>*SD* = 1.363 | 0.365 | 0.002* | 0.334 |
| Task relevant to all humans | *M* = 3.831<br>*SD* = 2.061 | *M* = 3.265<br>*SD* = 1.446 | 0.567 | 0.001* | 0.301 |
| Task relevant to me personally | *M* = 2.961<br>*SD* = 1.060 | *M* = 2.954<br>*SD* = 1.155 | 0.007 | 0.960 | 0.005 |

* Signifies Levine's test for equality of variances violated, equal variances not assumed test statistics used.
** Cohen's *d* effect size (maximum likelihood estimator).

we could not treat every participant as coming from the same baseline condition when completing the second survey. Thus, we created a difference score by subtracting scores in the first advisor evaluation from the second. We then conducted a $2 \times 2 \times 2$ ANOVA with task type, advisor type, and replacement type as factors. Results indicated no significant interaction between the three factors and advice appropriateness, $F(1,659) = 1.297$, $p = 0.255$, $d = 0.086$; but there was a significant interaction for advice usefulness, $F(1,662) = 5.829$, $p = 0.016$, $d = 0.189$, and advisor expertise $F(1,662) = 3.940$, $p = 0.048$, $d = 0.155$.

To understand these interactions, we conducted two-way ANOVAs comparing the effect of replacement and advisor type on expertise and advice usefulness within each task. In the humanitarian task there was a significant interaction between replacement and advisor type for advisor expertise, $F(1,355) = 13.156$, $p < 0.001$, $d = 0.288$; this interaction was not significant in the management task condition, $F(1,305) = 0.058$, $p = 0.810$, $d = 0.005$. The interaction analyses were similar for advice usefulness: significant in the humanitarian task, $F(1,360) = 8.205$, $p = 0.004$, $d = 0.0224$, but not in the management task, $F(1,300) = 0.605$, $p = 0.437$, $d = 0.051$. An inspection of the means indicates machine advisors replacing human advisors in the humanitarian task produced the largest decrease in evaluations of advisor expertise ($M$ = -0.439, $SD$ = 0.714) and advice usefulness ($M$ = -0.317, $SD$ = 0.627), whilst human advisors replacing machine advisors produced the largest increase in ratings of expertise ($M$ = 0.038, $SD$ = 0.818) and advice usefulness ($M$ = 0.063, $SD$ = 0.766). Thus, we find partial support for hypothesis 1a and 1b, human advice is perceived as more expert and more useful than machine advice in humanitarian decision-making tasks, but only when one advisor type has replaced the other. A table detailing the three-factor ANOVAs for H1a–c, H3a–b, and our RQs can be found in Table 2; see figures for graphs of significant three-way interactions.

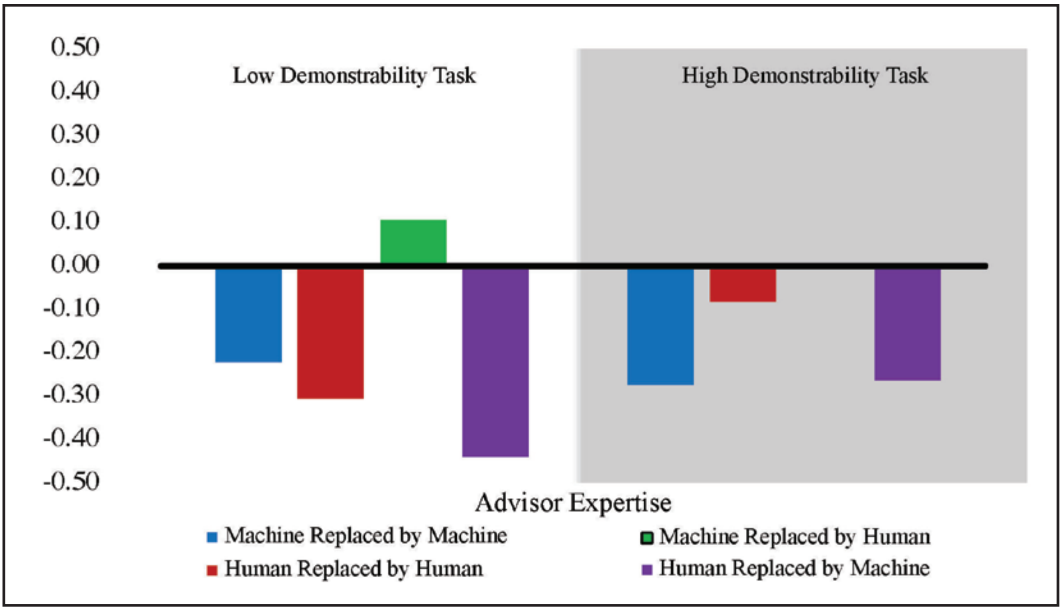



**FIGURE 1 Advisor Expertise Change from Advisor 1 to Advisor 2**

*Positive values = more perceived advisor expertise with second advisor*
*Negative values = less perceived advisor expertise with second advisor*

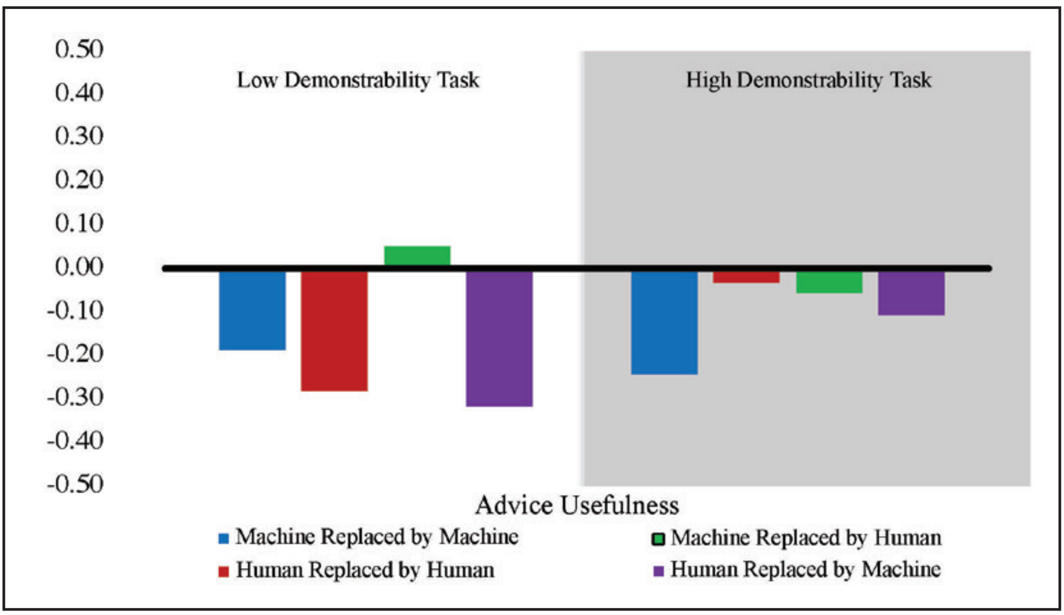



**FIGURE 2 Advice Usefulness Change from Advisor 1 to Advisor 2**

*Positive values = more perceived advice usefulness with second advisor*
*Negative values = less perceived advice usefulness with second advisor*

Hypothesis 2 stated that in the less demonstrable task (humanitarian), machine advice would be used less, in general, than human advice. Because all participants started the first 10 trials with a human or machine advisor, we first conducted a 2 (humanitarian/management task) × 2 (machine/human advisor) ANOVA with average advice utilization as the dependent variable for the first 10 trials. There was no significant interaction between advisor and task type $F(1,671) = 3.136$, $p = 0.077$, $\eta p^2 = 0.005$. There was a main effect of task ($F(1,671) = 8.775$, $p = 0.002$, $\eta p^2 = 0.013$) that indicated advice was used significantly more in the humanitarian ($M = 0.561$, $SD = 0.191$) than management task ($M = 0.521$, $SD = 0.188$), see Figure 1 in the appendix. To analyze the second block of trials, we had to account for similar or different advisor replacement as well as advisor type and task. We computed a variable composed of average advice utilization on the second block of trials and then conducted a 2(human/machine advisor) × 2(similar/different advisor replacement) × 2(humanitarian/management task) ANOVA but found no significant interaction between task, advisor, and replacement type, $F(1,667) = 0.656$, $p = 0.418$. We conducted a follow-up 2(humanitarian/management task) × 2(machine/human advisor) ANOVA but we did not observe a significant interaction, $F(1,667) = 0.320$, $p = 0.858$, or observe a main effect of task ($p = 0.133$) or advisor ($p = 0.602$). In sum, hypothesis 2 is not supported.

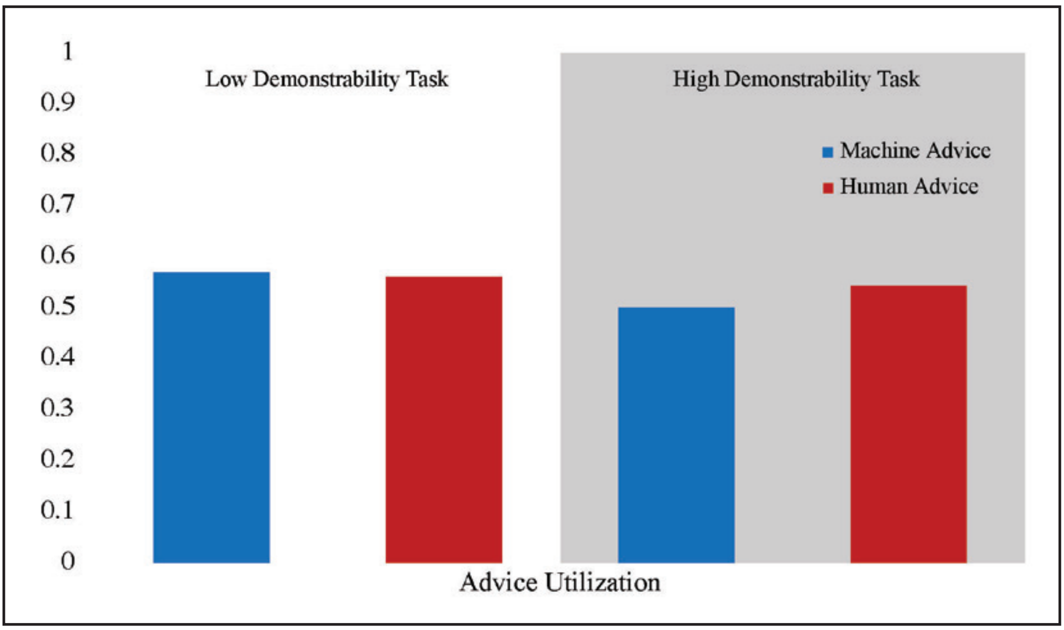



**FIGURE 3 Advice Utilization First 10 Trials**

*Higher values = more advice utilization*

Hypothesis 3a and 3b stated that perceptions of human/machine advisor thought process and value similarity would be greater/less in the humanitarian compared to the management task. For brevity, we used similar analyses to hypothesis 1, but with perceptions of value and thought process similarity. In the first block of trials there was no significant interaction between advisor and task type for perceptions of thought process similarity, $F(1,662) = 0.109$, $p = 0.741$, or value similarity, $F(1,656) = 0.256$, $p = 0.613$. Tests on the second block of trials revealed a significant three-way interaction of task, advisor, and replacement type on perceptions of thought process similarity $F(1,645) = 7.067$, $p = 0.008$, $d = 0.208$; but not on perceptions of value similarity $F(1,634) = 1.236$, $p = 0.267$. Follow-up two-way ANOVAs indicated a significant interaction between advisor and replacement type in the humanitarian task condition $F(1,352) = 38.038$, $p < 0.001$ $d = 0.672$, but in the management task condition this interaction was only marginally significant, $F(1,297) = 3.556$, $p = 0.060$, $d = 0.217$. Similar to ratings of advice usefulness (H1b), a human advisor replacing a machine advisor resulted in an increase in perceptions of thought process similarity ($M = 0.464$, $SD = 1.282$), whereas a machine advisor replacing a human resulted in the largest decrease ($M = -1.156$, $SD = 1.298$). In sum, we find partial support for hypothesis 3a, decision-makers do perceive more thought process similarity with human advisors compared to machine advisors in humanitarian decision-making scenarios, but only when one advisor type has replaced the other. We did not find support for Hypothesis 3b regarding value similarity; results are summarized in Table 2.

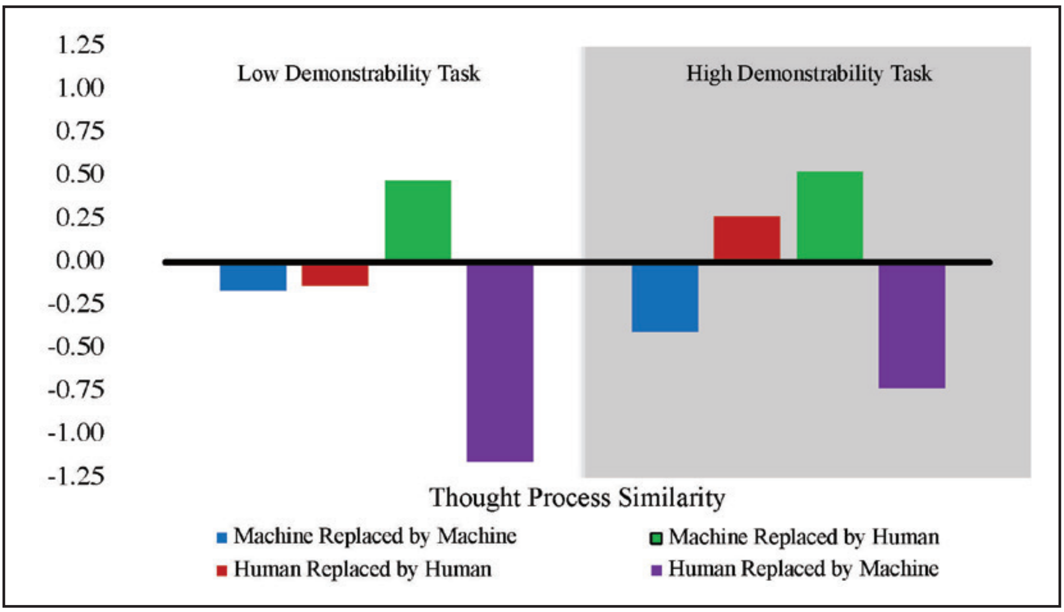



**FIGURE 4 Thought Process Similarity Change from Advisor 1 to Advisor 2**

*Positive values = more perceived similarity with second advisor*
*Negative values = less perceived similarity with second advisor*

**TABLE 2 Difference Between Rating of First Advisor and Second Advisor, Three-Way ANOVAs and Follow-Up**

| Measure | Interaction | $F$ | $p$ | $\eta p^{2*}$ |
|---|---|---|---|---|
| **Expertise (H1a)[1,2]** | **Task*Replacement*2nd Advisor** | **3.954** | **0.047** | **0.006** |
| **- In Humanitarian Task** | **Replacement*2nd Advisor** | **13.966** | **0.001** | **0.037** |
| - In Management Task | Replacement*2nd Advisor | 0.100 | 0.752 | 0.001 |
| **Usefulness (H1b)** | **Task*Replacement*2nd Advisor** | **5.494** | **0.019** | **0.008** |
| **- In Humanitarian Task** | **Replacement*2nd Advisor** | **7.786** | **0.006** | **0.021** |
| - In Management Task | Replacement*2nd Advisor | 0.565 | 0.453 | 0.002 |
| Appropriateness (H1c) | Task*Replacement*2nd Advisor | 1.381 | 0.241 | 0.002 |
| **Thought Process Similarity (H3a)** | **Task*Replacement*2nd Advisor** | **6.817** | **0.009** | **0.011** |
| **- In Humanitarian Task** | **Replacement*2nd Advisor** | **38.369** | **0.001** | **0.099** |
| **- In Management Task** | **Replacement*2nd Advisor** | **3.931** | **0.048** | **0.013** |
| Value Similarity (H3b) | Task*Replacement*2nd Advisor | 1.202 | 0.273 | 0.002 |
| Positive Emotions (RQ1) | Task*Replacement*2nd Advisor | 2.528 | 0.112 | 0.004 |
| **Negative Emotions (RQ1)** | **Task*Replacement*2nd Advisor** | **3.888** | **0.049** | **0.006** |
| **- In Humanitarian Task** | **Replacement*2nd Advisor** | **4.114** | **0.043** | **0.011** |
| - In Management Task | Replacement*2nd Advisor | 0.673 | 0.413 | 0.002 |
| Reciprocity (RQ2) | Task*Replacement*2nd Advisor | 0.976 | 0.324 | 0.001 |
| **Faulting the Advisor (RQ3)** | **Task*Replacement*2nd Advisor** | **9.659** | **0.002** | **0.014** |
| **- In Humanitarian Task** | **Replacement*2nd Advisor** | **12.434** | **0.001** | **0.034** |
| - In Management Task | Replacement*2nd Advisor | 0.814 | 0.368 | 0.003 |

[1] Significant three-way interaction graphs in Figure 2.2–5 (Appendix).
[2] **Bold** = significant at the $p = 0.05$ level.
* Effect size (partial eta-squared).

Hypothesis 4a–4d suggested that the effects posited by hypotheses above would be affected by advisor replacement, making the observed effects stronger. Our results above effectively answered these questions. We find partial support for hypothesis 4a and 4b (see results for H1a, H3a) that referred to perceptions advisor expertise, advice usefulness, advice appropriateness, and perceived advisor thought process similarity. Machine advisors that replace human advisors are rated as having less expertise, less useful advice, and having less similar thought processes to the decision-maker when they (machine advisors) replace human advisors compared to when they replace other machine advisors in the humanitarian task, but this interaction between advisor type and replacement effect is not present in the management task. Similarly, human advisors that replace machines are rated as having more expertise, advice usefulness, and similar thought processes to the decision-maker when they (human advisors) replace machines as opposed to replacing another human; again, this effect is present in humanitarian but not management tasks. We do not find that ratings of appropriateness are significantly affected by advisor replacement. Hypotheses 4c and 4d suggested there would be an effect of advisor replacement type on utilization as well, but there was no significant interaction between task, advisor, and replacement type on advice utilization (see results for H2), hypotheses 4e and 4f are unsupported.

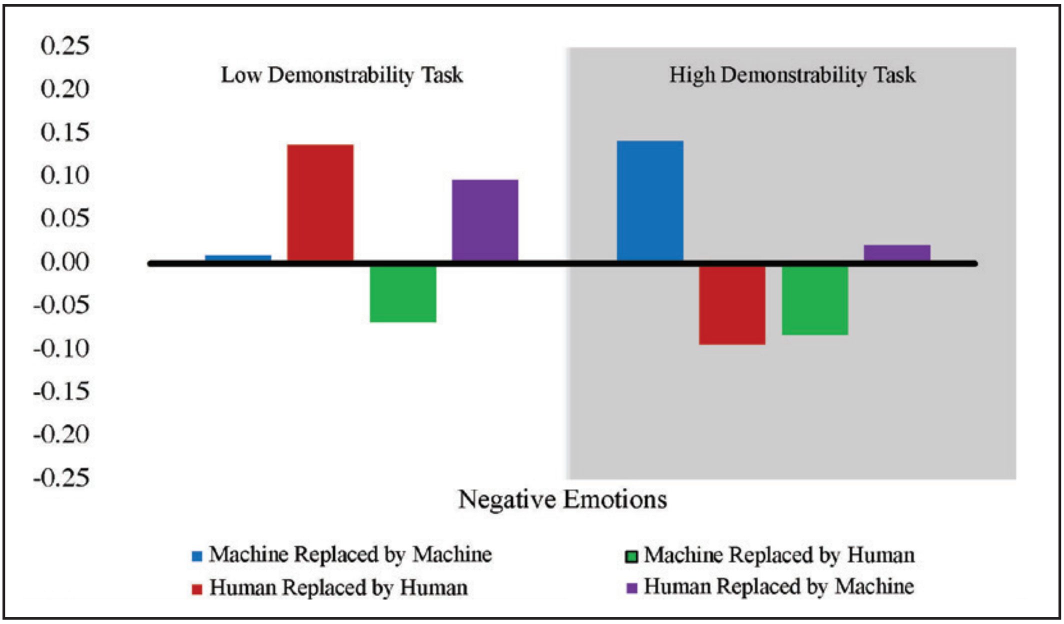



**FIGURE 5 Negative Emotions Change From Advisor 1 to Advisor 2**

*Positive values = more negative emotions with second advisor*
*Negative values = less negative emotions with second advisor*

With regard to RQ1 (emotions), RQ2 (reciprocity), RQ3 (fault), our analyses above suggested that the most interesting investigation would be in the difference between the rating of both advisors. We summarize the results of both exploratory paired samples *t*-tests and tests of three-way interactions between task, advisor, and replacement type in Table 3. Although these tests have not been corrected for type I error rate, we find them sufficient to answer our research questions that both decision-making task and advisor replacement type affect the emotions of decision-makers, reciprocity, and fault.

**TABLE 3 Summary of paired samples *t*-tests for RQ1-3. $M_{Diff}$ is mean difference of second advisor rating—first advisor rating (negative value = less endorsement of survey item for second advisor compared to first advisor).**

| Item | 1 | Humanitarian Task | | | | Management Task | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | $M_{Diff}$ | *t* | *Df* | *p* | $M_{Diff}$ | *t* | *df* | *p* |
| Positive Emotion | MrM | -0.09 | 1.05 | 76 | 0.306 | **-0.09**[2] | **2.78** | **74** | **0.001** |
| | MrH | 0.09 | 1.18 | 102 | 0.246 | -0.03 | 0.32 | 80 | 0.755 |
| | HrM | **-0.34** | **5.13** | **107** | **0.001** | -0.16 | 1.56 | 71 | 0.125 |
| | HrH | -0.15 | 1.41 | 71 | 0.161 | -0.03 | 0.30 | 73 | 0.762 |
| Negative Emotion* | MrM | 0.01 | 0.12 | 78 | 0.90 | **0.14** | **2.67** | **74** | **0.01** |
| | MrH | -0.05 | 0.78 | 105 | 0.441 | -0.07 | 0.81 | 82 | 0.422 |
| | HrM | **0.10** | **1.91** | **105** | **0.054** | 0.02 | 0.27 | 70 | 0.793 |
| | HrH | 0.14 | 1.31 | 70 | 0.198 | -0.09 | 1.08 | 74 | 0.292 |
| Reciprocity | MrM | **0.26** | **2.03** | **76** | **0.052** | 0.08 | 0.70 | 74 | 0.494 |
| | MrH | **0.74** | **5.02** | **102** | **0.001** | **0.89** | **5.08** | **80** | **0.001** |
| | HrM | **-0.96** | **7.62** | **105** | **0.001** | **-0.86** | **4.22** | **71** | **0.001** |
| | HrH | -0.12 | 0.95 | 71 | 0.354 | 0.20 | 1.36 | 75 | 0.183 |
| Fault the advisor for mistakes* | MrM | 0.19 | 1.43 | 78 | 0.160 | 0.13 | 1.12 | 76 | 0.278 |
| | MrH | -0.13 | 1.05 | 105 | 0.304 | -0.08 | 0.58 | 82 | 0.577 |
| | HrM | **0.70** | **4.82** | **106** | **0.001** | **0.35** | **2.16** | **72** | **0.035** |
| | HrH | **0.40** | **2.45** | **71** | **0.002** | -0.07 | 0.48 | 75 | 0.639 |

1: Advisor/Replacement Type: MrM = Machine Advisor replaced by Machine Advisor (similar replacement condition). MrH = Machine replaced by Human Advisor (different replacement condition), etc.
*Advisor Type × Replacement Type × Task Type three-way interaction is significant at the 0.05 level.
[2] **Bold** = significant at the $p = 0.05$ level.

## Discussion

Technological innovation is leading to the increased prevalence of algorithmic, machine advice in personal and professional life for decisions of varying demonstrability in fields as diverse as medicine, financial advising, and consumer goods. Additionally, machines are increasingly replacing human workers, and this trend is being exacerbated by recent events including the Covid-19 pandemic and advancements in artificial intelligence (Hayasaki, 2020). Our results show how the field of human-machine communication can use extant research from many communication subfields to inform our understanding of an increasingly automated world. To summarize, our findings show effects relating to the perception of an advisor as well as actual advice utilization, although both sets of effects are not always related. We found support for our first hypotheses which suggested human advice would be perceived as more expert and useful than machine advice for a humanitarian relief planning decision than a management decision. However, this effect was only significant when a machine advisor had replaced a human advisor and vice versa. Our third hypothesis suggested that advisors would also differ across tasks on decision-maker's perceptions of advisor similarity (humans perceived as more similar in humanitarian tasks). We found that human advisors were perceived as having more thought process similarity (to the decision-maker) but, again, only when the human advisor had replaced an advisor and vice versa. We did not find significant effects of task demonstrability or advisor replacement on perceptions of advice appropriateness or perceptions of advisor value similarity. There were no significant differences in advice-utilization once the decision-makers had their advisors replaced. Overall, there are not large effects of task and advisor type on utilization. Finally, our research questions showed that decision-maker emotions, reciprocity, and fault can be affected by advisor replacement and task type.

## Perception and Behavior With Machines

There are a number of potential explanations for why the manipulation of advisor type, replacement, and task type may affect *perceptions* of advisors differently than affecting actual utilization *behavior*. One explanation regards the difficulty and unfamiliarity of the task. Interpersonal advice research finds that decision-makers seek and utilize advice more when they perceive tasks to be more difficult (Bonaccio & Dalal, 2006). If decision-makers in our experiment perceived the actual act of forecasting as something they were not able to do well, it may have driven the utilization of advice regardless of perception of the advisor or perceived quality of the advice. In other words, although one advisor was perceived more positively, participants may have still perceived either advisor as more informed than themselves. An interesting manipulation for future study is to select easier tasks or tasks on which decision-makers perceive themselves to have expertise. People who believe they are experts are more prone to overconfidence and advice-discounting in general (Bonaccio & Dalal, 2006), and thus, there would be a higher bar toward advice utilization.

Another potential explanation relating to expertise is the possibility that differing ratings of expertise between advisors in task conditions was due to decision-makers feeling that human advisors had more expertise than machines in understanding the consequences of a decision (human lives), but not in actually comprehending the forecasting data and producing an optimal forecast. This is an interesting direction for future study in research comparing human versus machine advice because it may uncover further underlying assumptions that decision-makers have about machine and human advisors that are specific to different parts of a decision process (Einhorn & Hogarth, 1981). In our experiment, the graph stimuli remained exactly the same, but the numbers clearly meant something different (lives versus dollars). Thus, evaluation of the numbers themselves may involve a different cognitive process (i.e., information processing) than evaluating their meaning, which is a more judgmental process. Utilization of the advice may have reflected a decision-maker's assessment of advisor's ability to perform one aspect of the decision process, but perceptions of the advisor's expertise and usefulness may reflect an assessment of a different process such as judgment. Expectations of machines play a critical role in predicting detrimental behaviors such as over- or under-reliance on machines (Madhavan & Wiegmann, 2007), and a better understanding of what aspects of the decision-making process these expectations refer to may lead to better machine advisor design and integration.

## Emotion and Machines

Another potential moderator between perception and behavior is decision-maker emotions. We examined both positive and negative emotions due to past research showing emotion's effect on decision-maker perceptions and utilization (de Hooge et al., 2014). Our results showed a significant effect of task on negative emotions, but perhaps the more interesting result is that for positive emotions there was only one condition that produced an increase in positive emotions: when a human advisor replaced a machine advisor on humanitarian tasks. In the same task, machines replacing a human produced the largest decrease in positive emotions (see Table 3 for mean differences). These results should be interpreted with caution because they did not result in significant omnibus effects, but there is a clear direction implied—decision-makers are not feeling positive emotions when machines replace a human advisor when making a less demonstrable decision. In our study and interpersonal advice research in general, it is unclear how emotions are tied to utilization behavior. For example, Gino and Schweitzer (2008) found that inducing anxiety led to more advice utilization, but de Hooge et al. (2014) found that negative emotions resulted in lower perceived expertise of an advisor and lower utilization. Our results show that emotion is an important area for future research, especially because the emotional reaction to receiving advice seems to differ between human and machine advisors.

## Human Similarity and Liking

Perceived advisor similarity is another set of findings that provide insight into the complicated relationship between humans and machines. Advisor thought process similarity ratings for human advisors increased more when they replaced machine advisors. Although we do not know if decision-makers are consciously comparing one advisor to the other, a large amount of contrast effect research suggests this process happens unconsciously (Palmer & Gore, 2014). Perceived similarity does generally result in more liking (Strauss et al., 2010) and the implication of this is not only that humans may like human advisors that replace machines, but that humans do not like machines that replace other humans. This is an important area for future research; there is conflicting survey evidence about how much the average people like the idea of machines replacing humans (Savela et al., 2017), and some field research suggests that machines are sometimes welcomed as a replacement to humans (Wasen, 2010) or desired not to replace humans (Kristoffersson et al., 2011).

Human advisors also may be liked more in general because our results show that decision-makers feel more reciprocity (i.e., "I owe something to my advisor") toward human advisors. This is quite a remarkable result if one considers that our manipulation of human versus machine advisor was very minimal in this experiment. While agency and influence (Banks & de Graaf, 2020) were present for the advisor, there was almost no interactivity with either advisor nor was there any conversational wording added to the advice; it was simply delivered as a number. In conjunction with our results regarding advisor perception above, this result has important implications for human-machine trust theory—especially continued efforts to investigate what degree people see machines as social actors (Gambino et al., 2020). Our results suggest that even our small manipulation with no social interaction leads to very different assessments of a social feeling like reciprocity.

## Research Implications and Conclusion

Our results also have real-world implications. If a human is replaced by a machine in the workplace, the interpersonal advice process clearly is a more social process than the human advisor in this experiment. Yet, our experiment revealed this perception of a social process is substantially different for human versus machine advisors despite the advice being hardly social at all thus. In settings outside the lab, there are likely to be several differences between human and machine advice that would act as confounds if not controlled in a laboratory setting. Our experiment removed the confounds introduced by actual real-world social relationships between humans that exist in the workplace, and thus this was a very conservative comparison of humans versus machine. When humans with real relationships are replaced by machines, perhaps these elements of social interaction are not "replaced," but actually "lost" instead. Humans are social creatures and the feeling that someone is helping you is a good one. There could be serious long-term consequences to the lost positive emotions that come from social interaction, everything from organizational commitment to productivity (Oswald et al., 2015) is at risk when employees are not happy at work. Moving forward, gaining a more thorough understanding of what happens socially and emotionally when a human colleague is replaced by machines is critical.

Additional real-world implications of our study are numerous. It is clear that humans do not like it when machines replace a human advisor, even a human advisor who is zero-acquaintance and only imagined. Furthermore, our results suggest that decision-makers *really* do not like it when this replacement occurs on a task that is less demonstrable. But the negative feelings experienced when machines replace a human do not necessarily mean that the machine advisor will be used less than the human advisor. If anything, our utilization results suggested machine advisors were used more in the humanitarian task, the same task that produced the most negative evaluations of the machine advisor when it replaced a human. Understanding how the manipulation of advisor characteristics, situational context, decision-maker self-efficacy, and advice accuracy affect this complicated relationship is important if machine advisors are to be effectively introduced into the areas they are being developed for including health care, financial advising, and disaster management. These industries are just a few of the many which will see the increased presence of machine advisors—and this trend is only projected to increase as the Covid-19 pandemic has dramatically increased corporate efforts to automate workforces. In conclusion, our research shows the process of replacing human advisors with machines will be complicated. Moreover, our research shows that it is not only the humans who are replaced that will be unhappy; the people who must work with these new machines may not be happy either.

## Author Biographies

**Andrew Prahl** (PhD, University of Wisconsin-Madison) is an Assistant Professor at the Wee Kim Wee School of Communication & Information at Nanyang Technological University, Singapore. His research addresses the communication consequences of replacing humans with machines. More broadly, Andrew's research investigates the key issues raised by the automation of labor.

https://orcid.org/0000-0003-3675-3007

**Lyn M. Van Swol** (PhD, University of Illinois at Urbana-Champaign) is a professor at University of Wisconsin-Madison and associate editor of Group Dynamics. Her research examines utilization of advice and acceptance of information in contexts like automation, forestry and conservation behaviors, and hospital administration. She also researches group communication and group decision-making, especially focusing on factors that increase one's influence in a group.

https://orcid.org/0000-0002-2484-748X

## References

Abraham, F., Schmukler, S. L., & Tessada, J. (2019). *Robo-advisors: Investing through machines* (SSRN Scholarly Paper ID 3360125). World Bank Research and Policy Briefs No. 134881. https://papers.ssrn.com/abstract=3360125

Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J., & Rahwan, I. (2018). The moral machine experiment. *Nature*, *563*(7729), 59–64. https://doi.org/10.1038/s41586-018-0637-6

Banks, J., & de Graaf, M. (2020). Toward an agent-agnostic transmission model: Synthesizing anthropocentric and technocentric paradigms in communication. *Human-Machine Communication*, *1,* 19–36. https://doi.org/10.30658/hmc.1.2

Bonaccio, S., & Dalal, R. S. (2006). Advice taking and decision-making: An integrative literature review, and implications for the organizational sciences. *Organizational Behavior and Human Decision Processes*, *101*(2), 127–151. https://doi.org/10.1016/j.obhdp.2006.07.001

Bonaccio, S., & Van Swol, L. (2014). Combining information and judgments. In R. S. Dalal & S. Highhouse (Eds.), *Judgment and decision making at work* (pp. 178–198). Routledge.

CGPGrey. (2014, August 13). *Humans need not apply*. https://www.youtube.com/watch?v=7Pq-S557XQU

De Hooge, I. E., Verlegh, P. W. J., & Tzioti, S. C. (2014). Emotions in advice taking: The roles of agency and valence. *Journal of Behavioral Decision Making*, *27*(3), 246–258. https://doi.org/10.1002/bdm.1801

Deloitte. (2016). *The expansion of robo-advisory in wealth management*. Deloitte Deutschland Financial Services. https://www2.deloitte.com/de/de/pages/financial-services/articles/the-expansion-of-robo-advisory-in-wealth-management.html

De Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A. B., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied*, *22*(3), 331–349. https://doi.org/10.1037/xap0000092

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. https://doi.org/10.1037/xge0000033

Einhorn, H. J., & Hogarth, R. M. (1981). Behavioral decision theory: Processes of judgement and choice. *Annual Review of Psychology*, *32*(1), 53–88. https://doi.org/10.1146/annurev.ps.32.020181.000413

Fildes, R., & Goodwin, P. (2020). *Stability and innovation in the use of forecasting systems: A case study in a supply-chain company* (Department of Management Science Working Paper, Lancaster University 2020:1). https://doi.org/10.2139/ssrn.3548701

Fildes, R., Goodwin, P., & Lawrence, M. (2006). The design features of forecasting support systems and their effectiveness. *Decision Support Systems*, *42*(1), 351–361. https://doi.org/10.1016/j.dss.2005.01.003

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, *1,* 7–18. https://doi.org/10.30658/hmc.1.1

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger casa: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1,* 71–86. https://doi.org/10.30658/hmc.1.5

Gino, F., & Schweitzer, M. E. (2008). Blinded by anger or feeling the love: How emotions influence advice taking. *The Journal of Applied Psychology*, *93*(5), 1165–1173. https://doi.org/10.1037/0021-9010.93.5.1165

Guzman, A. (2020). Ontological boundaries between humans and computers and the implications for human-machine communication. *Human-Machine Communication*, *1*, 37–54. https://doi.org/10.30658/hmc.1.3

Hayasaki, E. (2020, June 17). Covid-19 could accelerate the robot takeover of human jobs. *MIT Technology Review*. https://www.technologyreview.com/2020/06/17/1003328/covid-19-could-accelerate-the-robot-takeover-of-human-jobs/

Hoff, K. A., & Bashir, M. (2015). Trust in automation integrating empirical evidence on factors that influence trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *57*(3), 407–434. https://doi.org/10.1177/0018720814547570

Horberg, E. J., Oveis, C., & Keltner, D. (2011). Emotions as moral amplifiers: An appraisal tendency approach to the influences of distinct emotions upon moral judgment. *Emotion Review*, *3*(3), 237–244. https://doi.org/10.1177/1754073911402384

Kahn, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., Gary, H. E., Reichert, A. L., Freier, N. G., & Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 33–40. https://doi.org/10.1145/2157689.2157696

Katz, J. E., & Halpern, D. (2014). Attitudes towards robots suitability for various jobs as affected robot appearance. *Behaviour & Information Technology*, *33*(9), 941–953. https://doi.org/10.1080/0144929X.2013.783115

Kristoffersson, A., Coradeschi, S., Loutfi, A., & Severinson-Eklundh, K. (2011). An exploratory study of health professionals' attitudes about robotic telepresence technology. *Journal of Technology in Human Services*, *29*(4), 263–283. https://doi.org/10.1080/15228835.2011.639509

Langlotz, C. P., Allen, B., Erickson, B. J., Kalpathy-Cramer, J., Bigelow, K., Cook, T. S., Flanders, A. E., Lungren, M. P., Mendelson, D. S., Rudie, J. D., Wang, G., & Kandarpa, K. (2019). A roadmap for foundational research on artificial intelligence in medical imaging. *Radiology*, *291*(3), 781–791. https://doi.org/10.1148/radiol.2019190613

Laughlin, P. R., & Ellis, A. L. (1986). Demonstrability and social combination processes on mathematical intellective tasks. *Journal of Experimental Social Psychology*, *22*(3), 177–189. https://doi.org/10.1016/0022-1031(86)90022-3

Lawrence, M., Goodwin, P., O'Connor, M., & Önkal, D. (2006). Judgmental forecasting: A review of progress over the last 25 years. *International Journal of Forecasting*, *22*(3), 493–518. https://doi.org/10.1016/j.ijforecast.2006.03.007

Levay, K. E., Freese, J., & Druckman, J. N. (2016). The demographic and political composition of mechanical turk samples. *SAGE Open*, *6*(1), 2158244016636433. https://doi.org/10.1177/2158244016636433

Lourenço, C. J. S., Dellaert, B. G. C., & Donkers, B. (2020). Whose algorithm says so: The relationships between type of firm, perceptions of trust and expertise, and the acceptance of financial robo-advice. *Journal of Interactive Marketing*, *49*, 107–124. https://doi.org/10.1016/j.intmar.2019.10.003

Lutz, C., & Tamò-Larrieux, A. (2020). The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots. *Human-Machine Communication*, *1*, 87–111. https://doi.org/10.30658/hmc.1.6

MacGeorge, E. L., Guntzviller, L. M., Hanasono, L. K., & Feng, B. (2013). Testing advice response theory in interactions with friends. *Communication Research*, *43*(2), 211–231. https://doi.org/10.1177/0093650213510938

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *Academy of Management. The Academy of Management Review*, *20*(3), 709. https://doi.org/10.5465/amr.1995.9508080335

Merritt, S. M., Unnerstall, J. L., Lee, D., & Huber, K. (2015). Measuring individual differences in the perfect automation schema. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 0018720815581247. https://doi.org/10.1177/0018720815581247

Önkal, D., Goodwin, P., Thomson, M., Gönül, M., & Pollock, A. (2009). The relative influence of advice from human experts and statistical methods on forecast adjustments. *Journal of Behavioral Decision Making*, *22*(4), 390–409. https://doi.org/10.1002/bdm.637

Oswald, A. J., Proto, E., & Sgroi, D. (2015). Happiness and productivity. *Journal of Labor Economics*, *33*(4), 789–822. https://doi.org/10.1086/681096

Pagano, M. (2014, August 17). Humans need not apply: The future of jobs is robot-shaped. *The Independent*. http://www.independent.co.uk/voices/comment/humans-need-not-apply-the-future-of-jobs-is-robot-shaped-9673643.html

Palmer, J. K., & Gore, J. S. (2014). A theory of contrast effects in performance appraisal and social cognitive judgments. *Psychological Studies*, *59*(4), 323–336. https://doi.org/10.1007/s12646-014-0282-6

Prahl, A., & Van Swol, L. (2017). Understanding algorithm aversion: When is advice from automation discounted? *Journal of Forecasting*, *36*(6), 691–702. https://doi.org/10.1002/for.2464

Rice, S., & Geels, K. (2010). Using system-wide trust theory to make predictions about dependence on four diagnostic aids. *The Journal of General Psychology*, *137*(4), 362–375. https://doi.org/10.1080/00221309.2010.499397

Roggeveen, S. (2014, August 18). Humans need not apply: An economic horror movie. *The Interpreter*.http://www.lowyinterpreter.org/post/2014/08/18/Humans-need-not-apply-An-economic-horror-movie.aspx

Savela, N., Turja, T., & Oksanen, A. (2017). Social acceptance of robots in different occupational fields: A systematic literature review. *International Journal of Social Robotics*, 1–10. https://doi.org/10.1007/s12369-017-0452-5

Sniezek, J. A., & Van Swol, L. (2001). Trust, confidence, and expertise in a judge-advisor system. *Organizational Behavior and Human Decision Processes*, *84*(2), 288–307. https://doi.org/10.1006/obhd.2000.2926

Strauss, J., Barrick, M., & Connerley, M. (2010). An investigation of personality similarity effects (relational and perceived) on peer and supervisor ratings and the role of familiarity and liking. *Journal of Occupational and Organizational Psychology*, *74*(5), 637–657. https://doi.org/10.1348/096317901167569

Takayama, L., Ju, W., & Nass, C. (2008). Beyond dirty, dangerous and dull: What everyday people think robots should do. *Proceedings of the 3rd ACM/IEEE International Conference on Human Robot Interaction*, 25–32. https://doi.org/10.1145/1349822.1349827

Tzioti, S. C., Wierenga, B., & Van Osselaer, S. M. J. (2014). The effect of intuitive advice justification on advice taking. *Journal of Behavioral Decision Making*, *27*(1), 66–77. https://doi.org/10.1002/bdm.1790

Van Swol, L. (2011). Forecasting another's enjoyment versus giving the right answer: Trust, shared values, task effects, and confidence in improving the acceptance of advice. *International Journal of Forecasting*, *27*(1), 103–120. https://doi.org/10.1016/j.ijforecast.2010.03.002

Wasen, K. (2010). Replacement of highly educated surgical assistants by robot technology in working life: Paradigm shift in the service sector. *International Journal of Social Robotics*, *2*(4), 431–438. https://doi.org/10.1007/s12369-010-0062-y

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, *52*, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Wedell, D. H., Parducci, A., & Geiselman, R. E. (1987). A formal analysis of ratings of physical attractiveness: Successive contrast and simultaneous assimilation. *Journal of Experimental Social Psychology*, *23*(3), 230–249. https://doi.org/10.1016/0022-1031(87)90034-5

Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, *170*, 105256. https://doi.org/10.1016/j.compag.2020.105256

# Appendix

1: Methods Information
2: Power Analysis
3: Task Screenshots

## 1: Additional Methods Information

***Design Considerations.*** Our literature review revealed the need for careful design. For example, research on how the perceived machine suitability for societal roles is affected by anthropomorphism has important design implications. First, we do not anthropomorphize the machine in order to provide the cleanest manipulation of advisor type. Second, we limit the social aspects of the advice exchange process; there is no direct interaction with either advisor type, and advice is delivered in a simple text format. Additional design implications are based on above review of advisor expertise—we avoid creating implied expertise by clearly introducing the human and machine advisors as having equivalent expertise. We also precisely control the accuracy of advice to rule out the confound of an advisor actually being better at a decision-making task. Our design therefore is optimized to discover differing assumptions that people have about the attributes of machines versus humans on tasks of different demonstrability.

***Survey Measures.*** The questionnaire measures were identical and measured perceptions of advice usefulness and advisor quality on a sematic differential scale. Additionally, Likert survey questions measured emotions when receiving advice, trust of advisor, similarity (value, social norm, and thought process) to advisor, and perceptions of advisor effort.

Positive emotions were measured with four Likert questions on a 1 (not at all) to 5 (extremely) scale for four positive emotions: Appreciative, Happy, Grateful, Thankful. The negative emotion scale was composed of Mad, Frustrated, Annoyed, Irritated. The four positive and negative emotion questions produced sufficient reliability (positive: $\alpha = 0.940$, negative: $\alpha = 0.943$), and the mean was used as an index of positive/negative emotion. Four sematic differential questions were used to measure advice usefulness (e.g., thoughtful, useful); and achieved sufficient reliability ($\alpha = 0.811$) and is hereon presented as an index of advice usefulness. Finally, in order to keep the survey a reasonable length, a pair of questions was asked to assess feelings of reciprocity to the advisor (i.e., “I feel like I owe something to my advisor for their help”).

***Advisor Descriptions.*** We pilot-tested 20 descriptions (10 human, 10 machines) with 23 undergraduate students and selected the descriptions that were closest to one another in ratings of perceived expertise, clarity, and performance expectancy.

*Intellective (high demonstrability) Task: Machine.* Your advisor today is a computer program called OptiLytics. OptiLytics is a software program used by the Gain Healthcare System to help with forecasting. The statistical models in OptiLytics have been built using 10 years of past Gain Healthcare data, as well as some data from the Center for Disease Control in the United States.

*Intellective (high demonstrability) Task: Human.* Your advisor today is Logan Girard. Logan is a medical doctor who has been working for Gain Healthcare for 10 years doing operating room management and hospital operations. Prior to joining Gain, Logan gained experience in healthcare management with the Center for Disease Control in the United States.

*Judgmental (low demonstrability) Task: Machine.* Your adviser today is a computer program called ReliefLytics. ReliefLytics is a computer program used by the UNHCR to help with forecasting. The statistical models in ReliefLytics have been built using 10 years of past UN data, as well as some data from the Center for Disease Control in the United States.

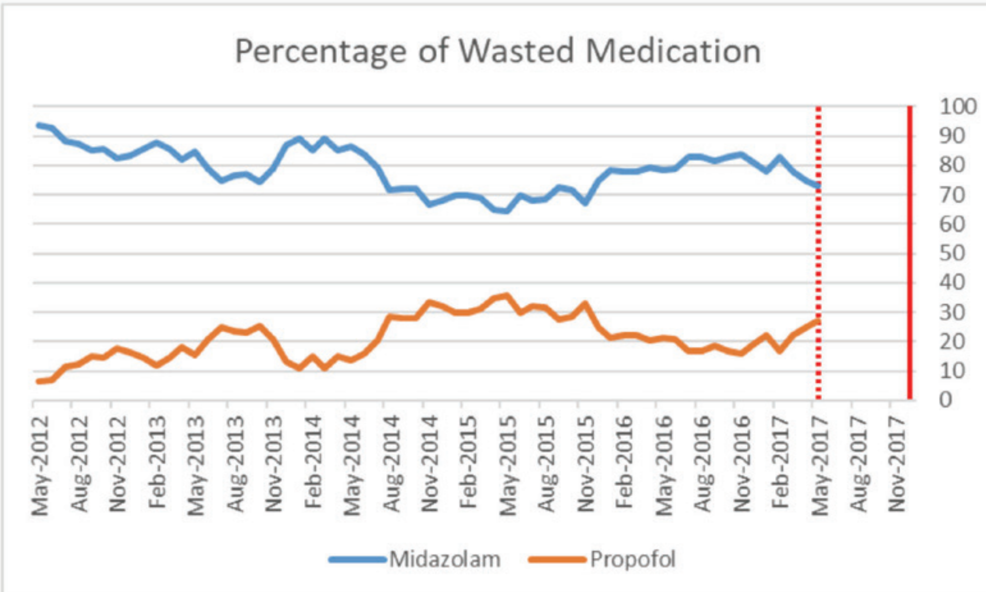
*Judgmental (low demonstrability) Task: Human.* Your adviser today is Logan Girard. Logan is a medical doctor who has been working for the UNHCR for 10 years doing camp management and emergency relief. Prior to joining the UNHCR, Logan gained experience managing medical crises in developing nations while working for the Center for Disease Control in the United States.

## 2: Power Analysis

A power analysis for the comparison between human and machine advisors was conducted using G*Power 3.1 and drew upon the three studies determined to be most similar to the research proposed (Dietvorst et al., 2015; Önkal et al., 2009; Prahl & Van Swol, 2017). Although these studies did not all report repeated measures results, the effect sizes were calculated as best as possible using published data. Dietvorst et al., reported effect sizes of 0.52 (Study 1) and 0.55 (Study 2); Önkal et al. effect size was calculated at 0.82; and Prahl & Van Swol reported a Cohen's d of 0.42. Of these, the most conservative estimate of total sample size needed by using the Prahl & Van Swol study, with G*Power calculating a needed N = 142 at alpha = 0.95 and a desired power of 0.80. Due to the well-known tendency of forecasting studies to experience high subject attrition due to missing data or the drawbacks of the weight of advice measure (for review, see Bonaccio & Dalal, 2006; Prahl & Van Swol, 2017; Tzioti et al., 2014) the target n is 162 in each advisor/task condition (human/machine & high/low demonstrability), leading to a total N = 648. Given the lack of previous studies, we have no power analyses for the advisor replacement effects, but subjects will be split into replacement conditions in each advisor/task condition and, given equivalent effect sizes, the above sample should be adequate.

## 3.1: Task Screenshots 1: Initial Forecast

Propofol and Midazolam are common drugs used to sedate patients, but often too much is prepared for the surgery and not used, resulting in us having to throw away some very expensive materials! Management wants to order less of sedative drugs in December 2017 to reduce waste.




Please help us plan this reduction: what percent of wasted sedative drugs that you think would be from Midazolam (blue line) in December 2017?

## 3.2: Task Screenshots 2: Feedback Screen

The correct forecast was: 72.8

The advice was 71.3
Advice forecast percent error = **2.1%**

Your revised forecast was: 71
Your forecast percent error = **2.5%**

Your average percentage error across all forecasts is currently:

**1.7 %**

This forecasting error costs the hospital: **$2500**

Your combined errors thus far are estimated to have cost the hospital: $3400

→