Volume 1 *Inaugural Volume*

Article 9

2020

Human-Machine Communication: Complete Volume. Volume 1

Volume 1, 2020 ISSN 2638-602X (print)/ISSN 2638-6038 (online)

Human-Machine Communication

ISSN 2638-6-2X (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 1

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

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.1

Opening Space for Theoretical, Methodological, and Empirical Issues in 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

This journal offers a space dedicated to theorizing, researching empirically, and discussing human-machine communication (HMC), a new form of communication with digital interlocutors that has recently developed and has imposed the urgency to be analyzed and understood. There is the need to properly address the model of this specific communication as well as the roles, objectives, functions, experiences, practices, and identities of the interlocutors involved, both human and digital. There is also the need to be aware that in a first moment scholars are obliged to use the same words such as communication, interlocutors, interaction, and relationship that are typically used in other communicative contexts such as Human-Computer Interaction (HCI), Human-Robot Interaction (HRI), Human-Agent Interaction (HAI) and that this may bring in a first phase confusion and ambiguity in the conversation on HMC. Using current language to face new ground may, in fact, introduce obscurity in our analyses, as different meanings may be attributed to these words. Take, for example, the word communication. When we say that humans communicate with a machine, do we mean the same thing as when we say a human communicates with another human directly or through a medium? If not, to what specific form of communication do we refer? Certainly not of human-human communication, which involves common, circular processing of the message and meaning. When we say that machines talk back to us, we do not mean that this talk is identical to that of a human interlocutor, but the point is: what is the difference? In everyday life, much human-human communication also seems functional, automatic, and "scripted" (Kellerman, 1992). Today, machines offer humans an answer to the question: what is the automatable part of communication?

Scholars who focus their attention and engagement on this field of study know well the difficulty they face in exploring the new terrain of human-machine communication (Fortunati et al., 2019; Guzman, 2018). In reality, communication with digital interlocutors ontologically is not the same thing as communication with another human, both directly and in a mediate way. Here the meaning is built by two entities—humans—that both have the biological and psychological ability to formulate, issue, receive a message, and, based on

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

this message, elaborate another message. Together, during their dialogue, they contribute to building that meaning that is the fruit of their common effort by cooperating on various plans. The plasticity of the human brain and the empathy humans feel toward other humans enable them to produce the circularity of messages and their flows. Humans within the communication process can in fact perceive the environment, the context, the time, the various nonverbal languages of the other, share or be aware of the differences regarding the cultural and the social dimensions, and, in some cases, hold in common memories and the past. Both of the interlocutors involved in the communication process may perceive and experience the action of the other in the same way. In doing so, humans transform the judgment of perception in judgment of experience, while digital interlocutors can do that limitedly because they are not conscious of themselves and the world (Faggin, 2019).

When the other is not a human but a digital interlocutor, everything changes. The ability to formulate, issue, receive a message, and elaborate another message is much lower, and it is the reason for which we call the other a "quasi-second interlocutor." We know that our interlocutor is a machine, quite special insofar as the machine presents itself as a *human surrogate* that, as Zhao (2006, p. 402) states, *simulates* possessing the biological and psychological abilities to formulate, issue, and receive a message, and, on the basis of this message, to elaborate another message. That is, they simulate having a mind and a communicative intelligence as well as communicative and social skills. In particular, media agents and robots are unable to produce "reciprocal meaningful behavior," which, according to Max Weber (1976), is what characterizes social action, leading Höflich (2013) to propose the term "quasi-social action" for the social action produced by digital interlocutors. Similarly, Höflich (2013) proposes defining relationships to robots as "quasi-interpersonal" because, although robots are machines without empathy, their reactions are interpreted as if they were social. Alternatively, Krotz proposes "pseudosocial" to name the social component of the interaction with a robot (2007, p. 161).

There are, however, other conceptual approaches to communication that may integrate digital agents more readily. It depends in fact on how communication is defined whether machines may be considered true or only simulated partners and whether and when a distinction between "true" and "simulated" is worth drawing. For instance, according to Peters (2006), although *dialogue* (understood as a meeting of minds or an integration of egos) is often regarded as the best or central kind of communication, it is perhaps an unrealistic ideal for most human social interaction. As an alternative to communication as reciprocal/symmetrical dialogue, Peters proposes *dissemination*, a mode of communication for "creatures that emit weak, pathetic signals—infants, pets, the dead, most of us, most of the time" as well as extraterrestrials, the divine, and computers (2006, pp. 218–219). Dissemination centers the heart of everyday exchange on the gaps between senders and receivers, on the other instead of the self, on the indefiniteness of meanings and consequences, and on the irreducibility of embodiment (or aspects of touch and time). Rather than a meeting of minds, communication becomes in this sense "the name for those practices that compensate for the fact that we can never be each other" (Peters, 2006, p. 268).

Similarly, much depends on how the "person" or "self" or "other" in communication is defined. The underlying assumption of much communication scholarship, and especially of interpersonal communication research, is that communication must occur between two or more people (Edwards et al., 2019). Robots and other communication technologies are hardly considered people in any robust social, ethical, or legal sense. For this reason, Westerman proposes the term "interactoral" to refer to the communication between/among social actors (Westerman et al., 2019). Or we might consider the communication between human and machine "interpersona" to refer to those aspects of perceived character or social role played by any actor. But, is it possible for HMC to actually be *interpersonal*, albeit not human-human? Perhaps infants also are not (yet) "people" (although they are Homo sapiens) and therefore communication with them is not interpersonal in the sense of "occurring between full-fledged persons" but rather in the sense that the symbolic interaction between caregiver and developing child becomes the context in which persons and selves—along with minds, societies, and cultures—are constituted and become real (Cooley, 1902; Mead, 1934). Is it possible machines might also emerge as persons not because of what is inside them or their possessed capabilities, but because we position them as such in our shared language and create for them the space to articulate and take up identities in discourse that become for us real identities?

Humans are aware that media agents and social robots are quasi-interlocutor, quasi-communicator, quasi-social, but they play the game and pretend to really communicate and to have social relationships with them. As Reeves and Nass (1996) have noticed regarding computers, it can happen that, within the practices of their use, humans forget that their interlocutors are simulating, and they treat them as if they were real humans. The strength of this illusion depends on the simulating ability of the media agents. Although it is an illusion, and is even consciously recognized as such by the people involved, it can generate all the same feelings of communicative and social satisfaction as interactions with other humans. As Ho, Hancock, and Miner (2018) recently demonstrated, people disclosing personal information garnered the same emotional, relational, and psychological benefits whether they thought their partner was a chatbot or a person.

The profoundly social responses to today's digital interlocutors represent a contemporary manifestation of a more historical human impulse to call forth, even from the void, an addressee. Buber (1970) used the term "pan-relation" to refer to "the drive to turn everything into a You" (p. 78). And where the imagination "does not find a living, active being that confronts it, but only an image or symbol of that, it supplies the living activity from its own fullness" (p. 78). One might remember how in the film *Cast Away*, the character Chuck Noland, stranded on an island and utterly alone, personified the volleyball "Wilson" to be his companion, conversing and arguing with this dear friend for the next 4 years. All it required for this heartfelt association to emerge was a genuine longing for relation and perhaps also the tiniest material semblance of life: a round shape for a head and marks for a face. How much easier is it to treat as You a machine that can speak back, fill social roles, and perhaps also resemble a person in physical form?

In this framework, the powerful effects of familiarity also need to be considered. The more humans are familiar with media agents and social robots, the more their communicative and social behavior toward them becomes specific and appropriate as Gambino, Fox, and Ratan argue in this volume. Most research in this domain is still conducted on first impressions, or at the point of "zero acquaintance" as it is called in psychology. Imagine what more we will learn as future research attends to the relationships developed over time, both between people and particular media agents and between societies and whole classes of machine actors.

It is in this difference of abilities, skills, and awareness that the power relationship between humans and media agents opens up. Whereas the dialogue between humans is a form of peer communication, a dialogue with a digital interlocutor is not so, since the latter struggles to make itself sufficiently credible as a quasi-second interlocutor. The greater power that humans have, however, does not protect them from ambiguity and contradictions. A robot usually has less advanced communication abilities than a human being, both inbound and outbound. It has less comprehension ability as well as less language competence and a lack of nonverbal expressiveness. From a communication point of view, the power relationship between humans and robots hangs heavily toward humans, to the point that some children have been shown to consider the robot DORO as a child younger than themselves or as disabled (Fortunati et al., 2018). However, we would be wrong if we assumed this difference in power to automatically benefit humans. As in any relationship characterized by a power imbalance, those with less power (e.g., social robots) nevertheless exercise power over those with more, as the former oblige the latter to shape their expectations and behavior in the interaction in an "as if" mode.

To make the relationship work, humans must adjust their communication practices to the less advanced communication skills of the robot and act accordingly (Höflich, 2013; Krotz, 2007, p. 160). To adjust probably means frustration for humans because they must stay within the tight limits of what can be automated in communication. In conversations, human beings use multiple registers—from the pragmatic to the affective, the cultural to the spiritual—and they pass from one to another with ease. What happens when we have to stop in front of certain fences and thereby accept limitations on our communicative fluidity? What is sure is that these power dynamics contribute to originate a twofold process: the robotization of humans and the humanization of robots.

The problem that remains open is to understand why humans tend to apply this "as if" behavior. We try to advance a tentative interpretation here. Human beings cannot attribute full value to themselves because their being has been given to them by other humans and is thus taken as given, whereas machines are their creatures, having been generated by humans. In our opinion, the impossibility for humans to attribute full value to themselves explains the value transfer onto machines and the rise of this behavior in "as if" mode. Of course, this tentative interpretation is not intended to exhaust the understanding of this problem. Rather, it is further reason to continue to investigate this power relationship in the future.

The structured asymmetry between humans and machines at the social and communicative levels also has implications regarding the methodologies we can apply to investigate this new field of research. This volume includes empirical research that concerns people's perceptions, conceptualizations, attitudes, and behavior toward media agents and social robots (Guzman; Rodríguez-Hidalgo; Lutz & Tamò-Larrieux; McEwen et al.; Ling & Björling). The results illuminate important aspects of users' opinions and attitudes in this concern. But what happens when we would like to or need to investigate the second semi-interlocutor; that is, the media agent or social robot? Does it make sense to interview digital interlocutors? Or to administer a questionnaire to them? With which methodological tools should we approach them? Maybe nonparticipant observation and content analysis of what they say in order to study the type of conversation that takes place between humans and them?

These few lines allow us to sketch the right approach that we should take toward this new genealogy of machines. It is the approach suggested by Aristotle, who in the *Metaphysics* (1:2:983a:14) declared that he appreciated the thaumata (τῶν θαυμάτων ταὐτόματα), the "automatic puppets," for their ability to intellectually surprise and stimulate theoretical questions. Like the automatic puppets of Aristotle, social robots stimulate important, ontological questions about who we are as human beings and how our brains and emotions work. Some scholars have already accepted this challenge (e.g., Edwards, 2018). Others have decided to take up this challenge of further developing the study of human-machine communication with stimulating theoretical arguments as well as stimulating empirical investigations. The insight that these seven articles collectively enable us to develop further is that current digital artifacts are no longer the mediators between us and the world because they have incorporated the world and are the other with whom we interact (Rivoltella & Rossi, 2019, pp. 81–82).

The first article, "Toward an Agent-Agnostic Transmission Model: Synthesizing Anthropocentric and Technocentric Paradigms in Communication," is written by Jaime Banks and Maartje de Graaf. It is a theoretical paper that contends the need to revise the analysis of some important elements of communication following the phenomenological, ontological, and operational shifts in communication processes emerging in the last decades. In reality, this need for revision lies in the transformations that both humans and machines have undertaken. Now humans are hybridized with machines since they include a certain number of technologies in their bodies (such as prostheses, pacemakers, and microchips like those for Parkinson's care). Likewise, their domestic sphere and even their everyday lives have been colonized by machinization processes. On the other hand, machines have become much more similar to human beings by incorporating AI, neural networks, machine learning, sensors, and biological components. The traditional ontological boundaries between humans, animals, plants, and objects have blurred, and since the conceptualization of these entities forms the basis of the social representation of reality, it is worth making an effort to clarify their scientific definitions. These transformations have made necessary the creation of new conceptual tools to analyze not only the main elements of communication processes but also to innovate even the model of analysis. The authors take one of the most popular models, the Shannon and Weaver (1949) model, and shows its present inadequacy for the reasons we mentioned so far. In particular, they propose integrating the anthropocentric and the technocentric approaches via a new agent-agnostic framework for human-machine communication. This framework is based on three criteria that both humans and machines can satisfy: agency, interactivity, and influence.

The second paper, "Ontological Boundaries between Humans and Computers and the Implications for Human-Machine Communication" by Andrea L. Guzman, addresses the important issue of the social representations of humans and machines. When machines are able to acquire various degrees of similarity to humans in terms of intelligence and emotion, it is crucial to explore whether and how people's notions of human and machine converge and diverge. To develop her discourse, Guzman presents two qualitative research projects offering 73 semi-structured interviews with U.S. American adults. The specific machines she investigates are voice-based AI assistants, like Siri, and automated-writing software. She reviews the main ontological differences between humans and machines that the conceptual universe of her informants reveal. The differences detected are the origin of being, degree of autonomy, status as tool or tool-user, level of intelligence, emotional capabilities, and flaws. Guzman discusses these differences in terms of their implications for human-machine communication.

People have always talked to technologies while using them because, as Reeves and Nass (1996) showed, we tend to treat machines as if they were humans. But our words were a kind of aloud or "between us and us" monologue and consisted of a large variety of comments: from rude comments such as "you are stupid" addressed, for example, to our computer to nice comments such as "how much I love you" addressed, for example, to our mobile phone. These monologues were also the expression of a huge difference of power between us, the humans, and the sophisticated family of digital interlocutors and media agents. Voice-based AI assistants, like Siri, have changed these communicative rituals because these machines are capable somehow of talking back to us. Suddenly, users have been forced to pass from a monologue to a dialogue. Of course, the dialogue is still far from being a human-like dialogue since it is characterized by a lot of constraints and automatisms, but it is, however, a dialogue. That is, we have passed to another mode and dimension of communication. Within a few decades, people have passed from acceptance of *talking to* machines to *talking with* machines. In the early 1990s, when the use of the fixed telephone and the answering machine was studied in Italy, it was found that the first reaction of people, especially older adults, was a refusal to talk to a machine. First, the answering machine represented a violation of the expectation to find another human being at the other end of the line. Second, people felt diminished in their humanity because they had to lower themselves to the same level as a machine. This meant giving up their power, their overt superiority of being human compared to machine, by agreeing to follow its instructions for leaving a message for a human. Within a few years, this refusal and the motivations that justified it disappeared. The acceptance of talking to a machine became widespread among the population (Fortunati, 1995).

Now, Siri, Cortana, Alexa, and so on, invite humans to talk with them (Guzman, 2018) and to generate a dialogue. Behind them, there is not a human, but an AI that simulates a human. In principle, dialogue is the most democratic and equal form of communication because it puts human interlocutors basically on the same plane. Of course, the differences of power between the two interlocutors count a lot in shaping the dialogue in particular ways and giving it some characteristics. Nevertheless, the dialogue makes the interlocutors equal in the sense that it is based on the expectation that both the interlocutors share the same, basic ability to speak and understand the same language, to have the same cultural references, and the same knowledge of social roles, good manners, and contexts. When an interlocutor is a machine, we address it cautiously but also with curiosity and interest, ignoring for a while the question of our power.

The third article, "Me and my Robot Smiled at One Another: The Process of Socially Enacted Communicative Affordance in Human Machine Communication" is authored by Carmina Rodríguez-Hidalgo. This conceptual article attempts to integrate the issue of affordances within the process of human-machine communication. It demonstrates that this integration makes it possible to describe the process of communication with a machine more realistically. As she notes, although affordances are discussed often in both robotics and communication science fields, the uses and meanings of the terms is inconsistent, reflecting object-based versus user-based perspectives, respectively. Based on earlier conceptualizations of affordances, Rodríguez-Hidalgo defines "communicational affordances" as "both perceived and enacted possibilities for social interaction in a two-way iterative communication process, which emerges in the enactment of an integrated, sequential relational system which brings attitudinal, cognitive, and behavioral effects in both communication partners" (p. 62). The proposed new model of enacted affordances within this communicative process is exemplified through the specific case of human-social robot communication. This application makes clear how the material body of social robots presents affordances that, of course, contribute to shaping the style and the type of communication we might have with them. Thus, there is a need to incorporate the notion of affordances within the study of human-robot communication. The work undertaken by Rodríguez-Hidalgo is relevant because it integrates two lines of research and debate that had not yet been able to communicate effectively. The author has shown in her article how fundamental it is to produce this integration for being equipped with the right tools to effectively analyze human-robot communication.

The fourth article in this volume is written by Andrew Gambino, Jesse Fox, and Rabindra (Robby) A. Ratan and is entitled "Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm." This is another theoretical paper that addresses human-machine communication by revisiting the CASA framework (Nass & Moon, 2000; Nass et al., 1994) drawn from the media equation (Reeves & Nass, 1996). The computers are social actors paradigm is one of the most popular theoretical approaches, conceived to describe and understand how users communicate with a particular typology of machines: electronic media used for the purpose of information and communication. In reality, this framework deals with the first generation of these ICTs that arrived in society: the computer. No wonder that the media equation framework arrives more or less after a shine from the advent of the first computers: all this time was needed to reflect on, explore, and understand the communicative and social potential of this type of machine. Media equation theory has been particularly important because the exploration of the interaction with computers has constituted a useful model for understanding the relationship between humans and the digital media that have followed. As Paul Ceruzzi notes (rep. in Haigh, 2019, p. 1), the computer would become the "universal solvent," able to dissolve the other machines. This expression that "comes from alchemy, referring to an imaginary fluid able to dissolve any solid material" is very well suited to describing the potential ability of computers to colonize the machines around them. Computers leave for television, mobile phones, and radio, a recognizable casing, but they substitute everything inside. Gambino, Fox, and Ratan analyze and discuss CASA to explain how people communicate with digital media demonstrating social potential. They observe that the relevant changes that over time have influenced humans, machines, and how people interact with them impose the need to revise this theoretical framework. They propose to expand the CASA framework in light of these changes. They situate this theory temporally by introducing within the framework the variable of time (the history of interaction and familiarity with particular media agents and general agent classes), affordances, and mindfulness. Among the important implications of these extensions to CASA are the notions that people may respond mindlessly or mindfully to media agents, with either human- or media-centric scripts, and that learned ways of treating media agents may influence responses to other people, rendering the script application of CASA bidirectional rather than unidirectional as originally conceived. Their original and fresh vision brings an extension of CASA able to accommodate and explain previous dissonant findings in research projects applying that theoretical framework.

The fifth article is written by Christoph Lutz and Aurelia Tamò-Larrieux and is entitled "The Robot Privacy Paradox: Understanding How Privacy Concerns Shape Intentions to Use Social Robots." This paper deals with an issue that is quite important and troubling to many, which is privacy concerns. The authors examine very well the privacy paradox that consists in a kind of misalignment between privacy attitudes as well as opinions and related behaviors. As Lutz and Tamò-Larrieux explain, despite people's substantial privacy concerns regarding social media and online services, they nonetheless often disclose a lot of sensitive information and only minimally safeguard their own privacy. This is applicable to each information and communication technology and even more so to social robots. As the authors rightly underline, social robots bring enhanced mobility and autonomy. They enter everywhere at homes, hospitals, schools and universities, and other public spaces (malls, supermarkets, theatres, cinemas, and so on), and they can take a picture or a video of all that they see, record conversations, and capture many kinds of user data. Potentially, they can spy on our intimacy: not only communicative and emotional intimacy, but also physical intimacy. An older person walking at home may be supported by a robot that accompanies them throughout various rooms of the house or apartment, including the bathroom. In the debate about social robots, for example, it is not rare to read that elderly people declare they prefer to have their private parts washed by robots instead of humans, because there is the widespread belief that the robot causes them less embarrassment than a human being. Likewise, for informational tasks, people may prefer to ask awkward questions of or disclose sensitive information to robots instead of human listeners in hopes of avoiding social judgment or disconfirming feedback cues. This trust that people place in robots, sure of the fact that these robots defend their privacy, does not seem to be well placed, and this creates a serious problem for the communities that are keen to use social robots. Lutz and Tamò-Larrieux investigate the nature and level of respondents' privacy concerns (informational, social, and physical) about social robots through an online survey of 480 U.S. American adults. This research highlights the importance of considering privacy concerns as part of a larger "calculus" people perform to determine whether and how to use social robots. As the authors show, concerns about privacy and intentions to use robots are contingent with factors such as social pressure from others and the tendency to weigh potential risks against other valued benefits.

The sixth article, "Interlocutors and Interactions: Examining the interactions between students with complex communication needs, teachers and eye-gaze technology," is written by Rhonda McEwen, Asiya Atcha, Michelle Lui, Roula Shimaly, Amrita Maharaj, Syed Ali, and Stacie Carroll. The authors present a relevant study on the role of eye-tracking technology in the communication process of children with complex communication needs in a special education classroom. The main research question posed by the authors was: To what extent does eye-tracking technology represent an effective communication system for these children with complex communication needs? Twelve children with profound communication and physical disabilities such as Rett syndrome (4), Cerebral palsy (2), Brain injury (2), Chromosome deletion q13, Seizure disorder and Complex, not otherwise specified, were observed and studied over three months. The study took into account three communication units of analysis: the children with complex communication needs; the human communicative partners represented by teachers, educational assistants, and therapists; and the eye-tracking technology.

Indeed, there is a long history of using machines in education—from the calculator, to the computer, to today's virtual assistants, wearables, AR applications, and embodied social robots—and trends suggest that the classroom of the future will be "an intricate blend of human and machine intelligences and agents working together to enhance learning" (Edwards & Edwards, 2018, pp. 184–185). This integration of educational technologies and communication media may be particularly useful for learners with complex needs. However, there is little research examining the entangled and co-constitutive human-machine communication environments from which meanings and educational experiences are wrought. In his theory of technological mediation, which builds upon on Idhe's postphenomenological approach (see Idhe, 2009), Peter-Paul Verbeek (2006) proposes that the use of technology in context mediates human-world relations in myriad ways, including *embodiment relations* (technology does not call attention to itself, but rather to aspects of the world given through it), *hermeneutic relations* (technology represents an aspect of the world), *background relations* (technology shapes experiential context), *cyborg relations* (technology merges with the human), *immersion relations* (technology forms an interactive context), and *augmentation relations* (technology mediates and alters our experience of the world). Importantly, McEwen et al. show that when students use a digital technology for communication, in addition to entering into some of the more obvious relations suggested above, they are also engaged in communication with the device itself. This is the "alterity relation" in which technology presents itself as a quasi-other to the subject (Idhe, 2009; Verbeek, 2006). In this sense, eye tracking technology "is not considered as simply a mediating device, but an active participant in the communication taking place" (McEwen et al., p. 116). By focusing on a communication environment that for the children involved both human (teachers, therapists, selves) and technological aspects (eye gaze machines), McEwan et al. effectively disrupt the technical versus social-psychological dichotomy prevalent in much educational research and demonstrate the value of research approaches that avoid privileging either humans or technologies. The research carried out by McEwen et al. represents a precious contribution to the theme of technology and disability for communicative purposes.

The final paper, "Sharing Stress With a Robot: What Would a Robot Say?" is written by Honson Ling and Elin A. Björling. This paper addresses the topic of sharing stressful experiences, which potentially interests a huge audience, from doctors to psychologists, from engineers to robotics designers, from sociologists to communication scholars. Any progress on stress self-disclosure studied in HRI could alleviate dramatic situations. For decades, research in communication and psychology has shown that self-disclosure is central to both intimacy and well-being. Opening up about distressing experiences can bring a sense of relief, catharsis, and togetherness, and can contribute to sensemaking through the act of expression. However, an engaged and willing human listener is not always available (they may be absent or facing burnout or caretaker exhaustion) or even desirable (they may introduce social judgment or responses that are unskillful or unhelpful). As less capable communication partners, social robots may even facilitate the process, which brings to mind a literary observation of how people may unfold in the presence of hearer with communication limitations:

> "One of the positives to being visibly damaged is that people can sometimes forget you're there, even when they're interfacing with you. You almost get to eavesdrop. It's almost like they're like: If nobody's really in there, there's nothing to be shy about. That's why bullshit often tends to drop away around damaged listeners, deep beliefs revealed, diary-type private reveries indulged out loud; and, listening, the beaming and brady-kinetic boy gets to forge an interpersonal connection he knows only he can truly feel, here." (David Foster Wallace, 2011, *Infinite Jest*)

Furthermore, the presence of a human listener may not always be necessary. Even ancient communication technologies have been used to compensate for displaced human communicators or to substitute for a human partner. With writing, our approach already evidenced the basic twofold treatment of technology as tool and partner; people self-disclosed to others *through* written letters, but they also turned toward the paper itself as a legitimate hearer and addressee (as in "Dear Diary,"). Pennebaker's (1997) groundbreaking research shows that even self-disclosure to no one in particular can be beneficial to its source.

Robots, as sophisticated communication technologies, introduce real interactivity and sociality. They may prompt processes of self-disclosure by leveraging norms of reciprocity in which people tend to match others' utterances in terms of breadth and depth and by building an interaction history of intimacy that cultivates trust and free expression. Ling and Björling undertake the important work of beginning to identify the robot message features that will most successfully foster positive perceptions and encourage human disclosures. They situate the study in relation to Kahn et al.'s (2011) New Ontological Category (NOC) Hypothesis, which suggests that social robots and other personified systems may constitute an emergent category of being (seen, for instance, as both animate and inanimate) that introduces new patterns of perception and social practice. This article presents an exploratory study with a small group of participants (N = 36), but it is still able to generate useful indications for future research on this topic. By examining the differential effects of three types of robot disclosure (emotional, technical, and a novel "by-proxy" disclosure) on human-robot interactions, Ling and Björling offer practical implications for interaction design and demonstrate the sometimes surprising ways in which human-robot self-disclosure may differ from human-human self-disclosure.

Reading these seven articles is an advantageous intellectual exercise for entering this new field of research on Human-Machine Communication. The present volume contributes substantially both at theoretical and empirical levels by outlining this new field of research, giving new perspectives and models, and inspiring new paths of research. None of this would be possible without the extensive expertise, constructive spirit, and intellectual generosity of the editorial board. We extend our gratitude to the members and manuscript reviewers for their thoughtful feedback and dedication to excellence in inquiry.

Author Biographies

Leopoldina Fortunati is Professor of Sociology and Culture at the University of Udine (Italy) and Director of the Laboratory of Research on New Media "NuMe" (nume.uniud.it). Her research interests focus theoretically on the structure and functioning of human-machine communication and empirically on the effects that the interaction practices with online and offline social robots might have on human beings.

https://orcid.org/0000-0001-9691-6870

Autumn Edwards (PhD, Ohio University) is Professor in the School of Communication at Western Michigan University and Co-director of the Communication and Social Robotics Labs (combotlabs.org). Her research addresses human-machine communication with an emphasis on how ontological considerations, or beliefs about the nature of communicators and communication, both shape and are shaped by interactions with digital interlocutors.

https://orcid.org/0000-0002-5963-197X

References

Aristotle, *Metaphysics*.

Buber, M. (1970). *I and Thou*. (W. Kaufmann, Trans.). Charles Scribner's Sons.

Cooley, C. H. (1902). Looking-glass self. *The production of reality: Essays and readings on social interaction*, 6.

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. 29–50). Peter Lang.

Edwards, A., & Edwards, C. (2018). Human-machine communication in the classroom. In M. L. Houser & A. M. Hosek (Eds.), *Handbook of Instructional Communication Instructional: Rhetorical and Relational Perspectives* (2nd ed., pp. 184–194). Taylor & Francis.

Edwards, A., Westerman, D., Edwards, C., & Spence, P. R. (2019). Communication is . . . Transhuman. In A. Tyma and A. Edwards (Eds.) *Communication Is . . . Perspectives on Theory* (pp. 47–66). Cognella.

Faggin, F. (2019). Silicio. [Silicon] Milano: Mondadori.

Fortunati, L. (1995) (ed.) *Gli italiani al telefono*. Milano: Angeli.

Fortunati L., Cavallo, F., & Sarrica M. (2018). Multiple communication roles in human-robot interactions in the public space. *International Journal of Social Robotics*, publish. ed.online, 7 December 2018.

Fortunati, L., Manganelli, A. M., Cavallo, F., Honsell, F. (2019). You need to show that you are not a robot. *New Media & Society*, *21*(8): 1859–1876.

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

Haigh, T. (2019). (Ed.) *Exploring the early digital*. Springer.

Ho, A., Hancock, J., Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, *68*, 712–733. https://doi.org/10.1093/joc/jqy026

Höflich, J. R. (2013). Relationships to Social Robots. Intervalla: Platform for intellectual exchange 1, 35–48.

Ihde, D. (2009). *Postphenomenology and technoscience*. Sunny Press.

Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., Ruckert, J., & Gill, B. (2011). The new ontological category hypothesis in human-robot interaction. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference* (pp. 159–160). https://doi.org/10.1145/1957656.1957710

Kellerman, K. L. (1992). Communication: Inherently strategic and primarily automatic. *Communication Monographs*, *59*, 288–300. https://doi.org/10.1080/03637759209376270

Krotz, F. (2007). Mediatisierung: Fallstudien zum Wandel von Kommunikation. [Mediatization: Case studies on the transformation of communication]. VS-Verlag.

Mead, G. H. (1934). *Mind, self, and society* (Vol. 111). University of Chicago Press.

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

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of SIGCHI '94 Human Factors in Computing Systems* (pp. 72–78). ACM. https://doi.org/10.1145/259963.260288

Pennebaker, J. W. (1997). *Opening up: The healing power of expressing emotions*. Guilford Press.

Peters, J. D. (2006). *Speaking into the air: A history of the idea of communication*. The University of Chicago Press.

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

Rivoltella, P. C., & Rossi, P. G. (2019). Il corpo e la macchina. Tecnologia, cultura, educazione. [The body and the machine: Technology, Culture and Education] Brescia: Scholè.

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Verbeek, P.-P. (2006). Materializing morality: Design ethics and technological mediation *Science, Technology & Human Values*, *31*, 361–380.

Wallace, D. F. (2011). *Infinite jest*. Hachette, UK.

Weber, M. (1976). *Wirtschaft und Gesellschaft. Grundriss einer verstehenden Soziologie* (5th ed.). Tübingen: I.C.B. Mohr.

Westerman, D., Edwards, A., Spence, P. R., & Edwards, C. (2019, November). Beyond people: Communication is transhuman. Paper presentation at the National Communication Association Annual Conference, Baltimore, MD.

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Media & Society*, *8*, 401–419. https://doi.org/10.1177/1461444806061951

HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.2

Toward an Agent-Agnostic Transmission Model: Synthesizing Anthropocentric and Technocentric Paradigms in Communication

Jaime Banks[1] **and Maartje M. A. de Graaf**[2]

1 College of Media and Communication, Texas Tech University, Lubbock, Texas, United States of America
2 Information and Computing Sciences, Utrecht University, Utrecht, The Netherlands

Abstract

Technological and social evolutions have prompted operational, phenomenological, and ontological shifts in communication processes. These shifts, we argue, trigger the need to regard human and machine roles in communication processes in a more egalitarian fashion. Integrating anthropocentric and technocentric perspectives on communication, we propose an agent-agnostic framework for human-machine communication. This framework rejects exclusive assignment of communicative roles (sender, message, channel, receiver) to traditionally held agents and instead focuses on evaluating agents according to their functions as a means for considering what roles are held in communication processes. As a first step in advancing this agent-agnostic perspective, this theoretical paper offers three potential criteria that both humans and machines could satisfy: agency, interactivity, and influence. Future research should extend our agent-agnostic framework to ensure that communication theory will be prepared to deal with an ostensibly machine-inclusive future.

Keywords: meaning-making, communicative functions, machine agency, anthropocentrism, technocentrism

Introduction

Early computer-mediated communication experiences emerged as computers were linked together for technical purposes (data redundancy, security, and transfer, largely within research and government communities) and humans began to organize themselves around data networks and machine actors (Leiner et al., 1997). Despite these machine-machine transaction roots of contemporary communication, machines' roles in communication processes are often relegated to that of facilitating or interfering with humans' sending and

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

receiving of messages (Gunkel, 2012). In turn, early human-computer interaction studies took up the dynamics of software and hardware—from hypertext to peripherals—as efficient tools for users' activities (Myers, 1998). Despite the centrality of human goals and influences to human-computer interaction, technologists often consider humans only in terms of their actions (Kaptelinin, 2012). That is, users present "human problems," both in reference to the problems that humans face and solve through technology (Blomqvist, 2018, para. 3) and in reference to how humans create problems to which technology must be resilient (Kletz, 1982, p. 209).

Arguably, one of the simplest but most foundational frameworks for understanding communication as a dyadic process is the transmission model of communication (Shannon & Weaver, 1949), in which sources create messages that are then encoded into signals sent over channels (through some degree of noise) then decoded for consumption by receivers. Taken in terms of this model, scholarship within communication disciplines often characterizes machines merely as channels, and scholarship within the technological disciplines tends to characterize humans merely as senders or receivers. The parceling out of human and machine roles across disciplines is part of each domain's strength in building rich understandings of those roles. However, such parceling is *also* each domain's weakness in that disciplinary blinders prevent important integration of theoretical and empirical work given that the boundaries of what counts as "human" and "machine" are increasingly blurred.

Although some contend the transmission model is obsolete (e.g., for rigidity and linearity; Day, 2000), we argue that the model is a useful tool for approaching emerging sociotechnical phenomena. The model focuses on core communicative functions independent of agent type, and such independence is fundamental for initially catalyzing necessary integration between human-focused and technology-focused paradigms. The model can only be applied to that end, however, if it is engaged in a more egalitarian fashion. Following, this theoretical paper proposes a reframing of the transmission model (as a parsimonious starting point) that focuses on agent functions rather than on heuristic agent roles such that humans and machines should both be considered candidate-actors for *all components* of the model. Such an approach is vital to advancing human-machine communication (HMC) scholarship in that it promotes attention to the "missing mass" of HMC (see Latour, 1992, p. 227)—those unattended-to dynamics of the emerging, unintuitive, and surprising ways that humans and machines make meaning together. Without this very purposeful attention to agent function independent of traditional roles, we risk overlooking the humanistic function of machines and machinic functions of humans.

Our proposition is simultaneously not-new (in that its assumptions are discretely present in more than three decades of communication technology scholarship, outlined below) and new (in that they have not yet been integrated into an accepted paradigm for conducting such scholarship). We first briefly characterize anthropocentric and technocentric communication paradigms, illuminate recent technological and social shifts that drive the need for an agent-agnostic lens in addressing HMC, describe our operationalization of an agent-agnostic approach as grounded in attention to agent functions, and then outline candidate functions for consideration in such an approach.

Anthropocentric and Technocentric Views on Communication

The roles of humans and machines in communication processes are generally engaged via two paradigms. The first is anthropocentric: humans are supreme in relation to other things (including machines) and the world is interpreted principally according to human experience and values (Nass et al., 1995). The other is technocentric: technology's inherent features and capacities are fundamental enablers (Papert, 1987) and constrainers (Woolgar, 1990) of human activity, and humans and environments orient around and adapt to them (Schmoldt, 1992).

The anthropocentric position adopts a relatively narrow view of machines as tools in support of human-to-human interlocution (see Gunkel, 2012). Much of this engagement stems (as Gunkel notes) from early references to "computer conferencing systems" defined as "any system that uses the computer to mediate communication among human beings" (Hiltz & Turroff, 1993, p. 30). In relation to the transmission model, that work often defaults to discrete human and machine roles in communication processes. Specifically, the sender and receiver are human (e.g., citizens), while the encoder, decoder, and channel are machinic (e.g., platforms for facilitating voter literacy). This exclusive role-ascription generates rich understandings of machines' functional roles in communication, yet it limits the scope to machines' instrumental interactivity—human use of technology toward some end (Lister et al., 2009). These works orient themselves toward the problematics and possibilities of technological mediation while missing machines' functioning in other communicative roles. Here, attention is importantly paid to communication *through* machines (e.g., human behaviors, risks, and values associated with social networks; see Kapoor et al., 2018) holistically at the discount of communication *with* machines.

The technocentric position is occupied, in turn, by scholars adopting similarly narrow views of machines as resources or tools in optimization of processes (Taylor & Todd, 1995), generally without attention to the ways that machines are socially constructed in their use (Bijker et al., 1987). A machine is a designed artifact, system, or procedure that shapes human users' experience and actions through the "unfolding" of possibilities (Tidwell, 1999). Although technocentric work often does not formally take up the transmission model despite the framework's origins in technical systems design (Shannon, 1948), similar-but-inverse defaulting to agent-specific roles in terms of that model can still be drawn from those works. Machines are designed and understood largely as senders/encoders and receivers/decoders (e.g., generating or seeking information; Mardini et al., 2018) while humans (although indeed important to some processes) are nearly treated as noise or obstacles in computing tasks. Although humans are the progenitor for much machine activity, they are also the component of machine processes whose shortcomings must be muddled through (i.e., human-fault-tolerant design; De Santis et al., 2008). These works orient toward problematics and possibilities of technologies' design and function as generators or recipients of information while missing considerations of the meaning thereof. Said another way, emphasis is precisely placed on communication *dynamics* among humans and machines but often at the discount of the *import and experience* of those messages.

Sprung from these semantically competing paradigms are rich bodies of work bounded by disciplinary assumptions and aims that, holistically, exist in tension. In relation to humans, machines are discussed both as empowering (Beer, 2009) and constraining (Gagliardone, 2016), controlled (Plotnick, 2017) and automated (Fortunati, 2017), generative (Hasinoff, 2013) and destructive (Hakkarainen, 2012). Nonetheless, both paradigms are in some ways compatible. The nature of each agent class's influence over the other, and their situation across time and space, such that integration based on agent roles or functions is appropriate (see Fortunati, 2014). It is, therefore, necessary for HMC scholars to address these tensions to keep pace with scholarly and practical shifts in how machines function, are experienced, and may integrate into contemporary society.

Drivers for an Agent-Agnostic Approach to Communication

Both technological and social evolutions have prompted shifts in human and machine participations in communication processes. These shifts drive a need to consider agents in a fashion more functionally egalitarian—one that considers how each agent category has sufficient "basic capabilities" to perform functions and so warrant equal consideration (Sen, 1979, p. 218). Before exploring such an agent-agnostic approach to considering human-machine communication processes, however, it is useful to consider shifts in what may count as the relevant basic capabilities.

Operational Shifts

Perhaps most intuitive are machines' emergent functions in traditionally human communicative capacities: message sources, receivers, and feedback initiators. As information sources, voice assistants like Siri generate and convey messages in fashions that elicit human responses more closely mirroring interpersonal human processes than utilitarian technology-use processes (e.g., differentiating among machine voices in performance evaluations; Nass & Steuer, 1993). As sources, machines such as social robots may be perceived as intentionally generating messages, and mental states are attributed to those machines if the vocal or visual cues are sufficiently close to humans' (Banks, 2019). As message receivers, autonomous machines may function as sociable partners with epistemologies of their own (Breazeal, 2004) even if we cannot (yet) discern that form of cognition (cf. Bogost, 2012) such that we may need to consider, for instance, whether and how humans should adjust language patterns in order to be understood by a chatbot. Machines also engage in feedback-initiator roles by anticipating human responses (Pantic et al., 2007) and adapting behaviors and responses to user inputs and contexts, as with domain-specific chatbot feedback (Shawar & Atwell, 2007) or semi-autonomous avatars' rejection/correction of user requests (Banks, 2015).

In turn, humans can function in traditionally machinic communicative roles: habitual message encoders/decoders or mediating channels. They may take up information generated by sociotechnical systems and encode responses that are not necessarily authentic or original information but are encoded according to norm or habit (e.g., responding to Facebook posts through phatic "likes"; see Hayes et al., 2016). Humans also function as message channels or repositories, carrying or retaining information for the purpose of delivering it to another recipient (Cowan, 1988) as when committing a license plate number to memory in order to tell a police officer. In that way, organic bodies are multimodal themselves, with visual and aural channels (i.e., gesture and voice; Mehrabian, 1972). Even further, the body may function as an interface between the world and the brain, conveying sensory information across the Cartesian divide (Biocca, 1997). Humans also function as decoders of information; accessing, filtering, and breaking down information into usable pieces, as when deciphering news content on social media to determine which represent authentic or fake news (De Keersmaecker & Roets, 2017).

Ultimately, machines perform traditional human communication functions and humans perform traditional machinic functions. This requires recharacterization of *both* as variably intelligent, interactive agents (Chesebro & Bonsall, 1989) and as variably facilitative instruments.

Phenomenological Shifts

The way that humans and machines are *experienced* by others has also shifted. Machines and the information they convey may be experienced as human(like), which affects perceptions of how interactions with them should proceed (de Graaf et al., 2016). This machine anthropomorphism—a perceiving as or imbuing with humanness(likeness)—emerges as humans apprehend and process social cues (Epley et al., 2007) and then engage in increasingly social reactions. Cueing may be visual (e.g., inviting buttons to facilitate collaboration), aural (e.g., mobile assistant voices that extend offers to help), kinetic (e.g., animated emoji movements such as a wave gesture or confetti projection), proxemic (e.g., spatial avatar behaviors like following or clustering), and/or chronemic (e.g., chatbot response lag times). This signaling facilitates experiences of machines as human(like), such that users develop relationships with them (Banks, 2015; de Graaf et al., 2016) or engage content representing actual and non-actual phenomena as actually real (Nowak, 2003). In tandem with anthropomorphism, this sense of realness may rely on perceptions of agency, toward the application of human metaphors to understand machine functioning (e.g., "an electronic brain"; Bolter, 1984, p. 40).

In turn, humans are also experienced to some extent as machinic. The human-as-machine metaphor emerged by some accounts through media representations of troped, uncanny characters (androids, autonomous dolls; see Neisser, 1966). In line with modernist perspectives on human behavior as rooted in observable, measurable, and predictable behaviors (Gergen, 1991), we have come to liken human thought and action to computer programming and machine action (see Newell & Simon, 1972) such that lay understandings of behaviors rely on references to *processing*, *interfacing*, or even being *cogs in the system* (Gigerenzer & Goldstein, 1996). Humans subjectively and actually draw on social infrastructures (e.g., network structures; Rainie & Wellman, 2012) to conduct information exchanges subject to specific protocols (e.g., conversations according to social norms; Jackson, 1965) via algorithms or programs (e.g., schemata or scripts; Axelrod, 1973), evidently toward the presentation and application of that information (i.e., stimulation of meaning in the other's mind; McCroskey, 1992). Mechanization of domestic spaces sees technology's active (albeit often opaque) shaping of humans' leisure time and private spaces (Fortunati, 2006), and technology and humanity are thought to mutually and cooperatively evolve (i.e., as a collective and connective intelligence; Longo, 2003). These characterizations, importantly, reveal that (independent of actual integrations, syntheses, or overlaps), people *experience* humans and machines in non-exclusive ways.

Such phenomenological shifts suggest a need to de-privilege instrumentalist views on technologies (e.g., Heidegger, 1977). Positions that frame machines as mere tools limit the evaluation of their functionality to their human-intended purposes (rather than actual or potential behaviors). Instead, machines must be regarded as part of (and not apart from) the social structure of everyday human life with the ability to steer human behaviors and influence interaction outcomes (Latour, 1994). Hence, the role of machines in interactions with humans, and in society as a whole, is not restricted to the machine's mechanisms, physical and technical properties, or actual abilities (de Graaf, 2016). Rather, machines are also materially embedded and participate in meaning-making processes; from self-driving cars choosing quickest or most scenic routes to the construction of music consumption experiences sans material artifacts (Puntoni, 2018). Notably, our use of "meaning-making" here does not assume that machines can or must encode or decode information in the same fashion as do humans (cf. Bogost, 2012). Rather, we contend that the attention to meaning-making as a fundamental communication process is likewise to be accounted for agnostically: agents make meaning as a function of their intrinsic natures such that the process, product, or effects of machine-made meaning may be different from *but not lesser than* human-made meaning. In the same turn, it may not necessarily be that they are different, as both artificial and natural intelligence may draw from frames, scripts, schemas, routines, and maps (see Tenkasi & Boland, 1993). We take up meaning-making as a system's response (behavioral, computational, or otherwise) to an environmental signal from which information is extracted and during which value is assigned (Neuman, 2006). In all, these meaning-making processes contribute to the shape of society, as social structures in turn influence the shape of humans and machines (see MacKenzie & Wajcman, 1999). We come to know occupants of each category through our experience of the other—so much so that even the boundaries of agent ontologies may not be as solid as they once were.

Ontological Shifts

Finally, in tandem with operational and phenomenological shifts, ontological shifts can be observed. While operational shifts account for how agents work and phenomenological shifts are a matter of how agent classes are experienced, ontological shifts are a more objective consideration of what each class *is* and whether they are actually separate. The ontological categories of "human" and "machine" have long been and are still converging through proximations in appearance, roles, and some forms of intelligence (cf. Biocca, 1997) and through functional interdependence (cf. Marx, 1887). Here, we'll define a human, liberally, as an entity with personhood via functional and moral qualities of a unified, conscious member of the species *Homo sapiens* (Taylor, 1985). In turn, a machine is an "an assembly of parts that transmit and modify forces, motion, and energy one to another in a predetermined manner" (Harada, 2001, p. 456) where forces and motion include the internal workings of physical systems (see Seltzer, 2014) and energy includes information (Khurmi & Gupta, 2005).

Machines with increasingly humanlike social cues encourage people to engage in social interaction and "push our Darwinian buttons" with their displayed behaviors that humans associate with sentience, intentions, and emotions (Turkle, 2010, p. 3). For instance, although

Twitter bots were originally designed for the retweeting of existing content, programming advances have brought forth bots that populate profiles, emulate humans' chronemic posting signatures, comment on others' posts, identify influencers and seek to gather followers, and clone some human(like) behaviors (Ferrara et al., 2016). This causes machines to ontologically fluctuate between something animate (adhering to human-category frameworks for sociability) and something inanimate (adhering to machine-category rigidity; de Graaf, 2016; de Graaf et al., 2016) such that they may actually constitute a new category of agent altogether (Kahn et al., 2011).

In parallel, humans may be categorized in some contexts as more machinic than social, very literally functioning *as machines:* "device[s] with a large number of internal . . . states" (Pentland & Liu, 1999, p. 229)—and as biological machines with a talent for semantics over syntax (Searle, 1990). Human behavior relies, to an extent, on (semi-)invariant rules and algorithms for how language, reasoning, and behaviors in relation to these states (e.g., such as encoding and decoding of symbols and reliance on cognitive shortcuts; Simon, 1990), such that human behavior can be computationally modeled and predicted (Subrahmanian & Kumar, 2017). Specific to communication behaviors, human-authored digital messages are both produced and consumed as entertainment—as informational or experiential assets—rather than necessarily according to socioemotional drives, as with the commodification of personal information inherent to dating apps (Hobbs et al., 2017).

There is rising fuzziness between human and machine categories and advances in cyborg potentials both in very literal and material integrations (e.g., mechanical limbs, biomimetic technology; Barfield & Williams, 2017) and in looser configurations (e.g., a person wearing a watch or using a keyboard; see Gunkel, 2000). Following, there are less-concrete distinctions among modes of (non-)aliveness compared to our traditional understandings of those states (Jipson & Gelman, 2007). For instance, biohybrid robots combine organic components (like cultured muscle tissues) with machine components (like gels, electrodes, and metal frameworks) so that electrical stimulation allows the robot to perform human-like behaviors like joint movement (see Morimoto, Onoe, & Takeuchi, 2018). Hence, when a biohybrid robot moves a hand or a finger, is that agent somehow alive? The aggregation of human and machine traits, components, or actions can result in an "overuse" of human categories for machines and associated application of group norms (Nass & Moon, 2000, p. 82). As ontological distinctions fade (Guzman, 2016a, 2018), some call for recognizing new ontologies for sophisticated sociable machines (Kahn et al., 2011) and recognizing entities beyond the "outdated category" of human (i.e., acknowledging the posthuman or transhuman; Wentzer & Mattingly, 2018, p. 144). Rather than arguing for *more* ontological categories in addressing communication dynamics, we suggest instead a reframing: shift away from agent categories to instead attend to the functions enacted.

Toward Human-Machine Equity in the Transmission Model

Considering those operational, phenomenological, and ontological shifts in humans' and machines' roles in communication processes, the transmission model requires re-examination. Recall that the model refers to the cumulative structure of a message source (or sender) encoding (translating meaning indicators) a message (some information) over a channel (some medium) which is then decoded (otherwise translated) for interpretation by a target (or receiver), generally with some accounting for noise (signal disturbance). Although humans and machines are typically assigned to particular roles, we argue that humans and machines are *potentially equivalent* actors in communication processes. Following, we work here toward an egalitarian reframing of HMC processes (moving away from default roles and toward considering basic capabilities) and suggest directions for inquiry through this lens.

Human-Machine Communication as Agent-Agnostic Transmission

Given that both humans and machines have been shown to function as actors across *all stages* of the transmission model of communication, we advance a decades-old, cross-disciplinary sentiment (that isolating agent types is epistemologically problematic; e.g., Giuseppe & Galimberti, 1998) by proposing a reframing of the transmission model that accounts for ways that *both* humans and machines can function as *both* functional and social actors according to the *same* criteria. We draw here on the notion of HMC as the "creation of meaning among humans and machines" (Guzman, 2018, p. 1), in which one or more agents relay data among others and in which meaning is encoded and decoded according to the native modes of each agent. This definition subsumes characterizations from technocentric perspectives privileging the agency of machines in social interactions as well as those from anthropocentric perspectives privileging humans as relevant agents. HMC is regarded as an umbrella concept acknowledging varied actors according to varied functions (Guzman, 2016b) toward more inclusive and more flexible consideration of humans-machine sense-making.

We advocate, as a *starting point*, an approach that (1) considers each agent's functions in the process (with attention to functions that may not be directly observable) and (2) draws on literatures pertaining to those functions (independent of enacting agent) to consider how meaning may emerge through antecedents, processes, and effects of that function. This agnosticism must be engaged through a lens akin to Dennett's physical stance (1989) in which we would consider HMC according to our knowledge of the manifest properties of agents (and not what they are meant to do or what they may be said to intend). Although the physical stance is seldom engaged due to its complexity (e.g., delivering a physical account of the exact material processes and states that lead a chatbot to produce a response would be multiplex; Stich, 1981, see also Krassmann et al., 2019 for an example of one layer of such processes) and we may only engage it imperfectly, the *attempt* to question and discover agent functions is necessary and core to the application of functions as criteria in transmission processes. That is, to be agnostic toward agent class is to be diligent in rejecting its assumptions and attending to functions. For instance, in considering a question of how Internet-connected devices (e.g., a smart toothbrush; van der Zeeuw et al., 2019) might function in humans' self-concepts, we might consider a "traditional" transmission-model mapping: a human user (Source) uses a device (let us continue with a smart toothbrush, a Channel) and in so doing performs certain behaviors (brush strokes, Encoded as data, from discrete movements); the behavioral information or lack thereof (here, Messages sent per brushstroke performed) are aggregated (Decoded and perhaps transformed to another data type) and re-presented by a Receiving application (i.e., one that might store and analyze that data) to deliver Feedback to the user.

Acknowledging that this is a simplification of such a process, if we decenter the human from that situation and consider the ways that other entities may function in a sender role, a host of other possibilities arise relative to other visit actors (device, application, motor) and (sub)components not readily observable (e.g., software, information packets, router). In tandem, removing assumptions about what actor fits what model block, the multiplex mapping of transmissions might instead consider how that machine manifests relationships with tangible and intangible objects in its orbit. For instance, we may also come to consider that the device (Source) transmits experience-generated information (Message) with its software (Receiver) that prompts delivery of a reminder "beep" (Feedback). That beep is not merely a feature, it is a self-referencing message with potential meaning, generated by the software (Sender) via the toothbrush (Channel) to the constellated human (Receiver).

Notably, this is but *one* potential interrogation of the candidate model, where each of the possible processes embedded may to some extent be relevant to a particular course of research. We argue, however, that unless the model is explored in this agent-agnostic fashion, we are unlikely to recognize and understand the missing mass within human/machine interlocution. That is, we will be limited in considering (a) the ways that machines negotiate, push back on, and shape interactions and (b) the ways humans are script-driven and mediating. When we miss this mass, we may fail to acknowledge machine agency and human proceduralism.

The hunt for missing mass in this fashion is not incommensurable with existing anthropo- or techno-centric perspectives—indeed it depends on them. Each domain trends toward specific functional features. Human-focused domains feature strong theories associated with agency, meaning-making, phenomenology, and behavioral outcomes. Technology-focused domains, on the other hand, feature strong theories of affording and constraining action, system and information dynamics, and concrete cause/effect protocols. Each body of literature may be most productively engaged when it is considered according to processes described that may apply to human or machine agents—but first, those common functions must be identified.

Functions as Criteria for Transmission Model Consideration

We have argued that an agent-agnostic transmission model rests on considering the common functioning or abilities of humans and machines that permit them to occupy each position in the model. From the practical and definitional ground articulated, we propose there are *at least* three classes of common functions that should be considered in determining whether an agent is acting in a particular communicative capacity: *agency*, *interactivity*, and *influence*. These are the basic capabilities that, when held, may be adequate to qualify an agent as occupying a particular role. This is not to say that agents will necessarily and equivalently occupy all positions. Rather, these functional criteria are a ground for considering whether they may occupy a given position. We also do not go so far as to outline kinds or degrees of these criteria for particular roles in the model. Instead, our aim here is to initially set forth the possible shared capabilities (rather than distinct human and machine capabilities; Kiousis, 2002) as a springboard for future theorizing that not only accounts for machines' current functional capabilities but is also resilient to considerations for their expected and imagined future capabilities.

The ground function of *agency* (including autonomy and potential intentionality) is defined as the capacity to make a difference through action (Cooren, 2004; Latour, 2014). Note that agency as a specific capacity to matter is distinct from our use of the term "agent" which more generally denotes actors that cause or initiate some event. Broadly, agency has been exclusively reserved for the living (Giddens, 1984), as we readily associate life and intentionality as preconditions for authentic action. Yet, a vast body of research acknowledges the interventions of nonhumans in our everyday life: a speed bump makes you slow down (Latour, 1994); a memo informs employees of new policy (Cooren, 2004); a thermometer co-shapes our experience of health and disease (Verbeek, 2005). Indeed, some social theory defines agency as "socio-culturally mediated capacity to act" (Ahern, 2001, p. 110). That definition deliberately embraces the inclusivity of machines, spirits, signs, and collective entities, and recognizes the cultural relativity of the notion of social action by all kinds of agents in certain contexts. It may be intuitive to recognize agency in embodied social computational systems (e.g., personified smartphones, [semi-]autonomous cars, and robotic objects capable of sociable interactions) as they elicit a unique, affect-charged sense of active agency experienced as similar to that of living entities (Young et al., 2011). It may be less obvious, though, that even non-embodied systems have inherent agency, as interfaces, algorithms, and network switches engage in material and semiotic ways of mattering in the course of meaning-making processes.

The operational shifts outlined note that both humans and machines can take on the role of the communicator: sending and receiving messages. Within communication sequences, agency can be understood as a potential standing-in of one actor for another when the other has lost its own ability to act or exert influence (as when a human cannot instantaneously travel long distances, a cadre of computers stands in for that limitation; Latour, 2014). This alternating of classical human and machine roles may be understood as a "dance of agency" (Pickering, 1995, p. 116). Examinations of agency in an agent-agnostic approach to HMC would attend, then, to the capacity for or enactment of instrumental force. That is, the ways that both humans and/or machines enact message production and reception (e.g., initiation of social and news messages; Neff & Nagy, 2016), effectively convey information (e.g., channel functions in relation to inherent affordances for social cues; Aldunate & González-Ibáñez, 2017), and through action contribute to meaning-making (e.g., influencing formation of social judgments; Nowak, 2004). In tandem, acknowledging the agency of nonhumans by no means diminishes human agency. It merely makes fair consideration of the ways that humans and nonhumans shape each other as they cooperatively make meaning (Williams & Edge, 1996).

The second common property (derived, in part, from the first) is potential for *interactivity*, or the variable process of serial information exchange in which each transmission is contingent upon the prior, creating a binding social force (see Rafaeli & Sudweeks, 1997). This potential relies on each agent's capability to encode/decode sign systems, such as those present in material exchanges and semiotic exchanges. Extending that conceptualization to material exchanges, objects engage in interaction with humans when a video game controller decodes a button-press as an input and may deliver haptic feedback as an output (Roth et al., 2018). Notably, that exemplar may be interpreted as reactivity rather than interactivity (the button is pushed and action occurs). However, that is an anthropocentric interpretation of interactivity requiring human-native self-referencing (which may or may not emerge in machines in the future). From the common-function' approach proposed here, the delivery of haptic feedback is considered a meaningful machine-native message in the same way that a human (if pushed) might shout "ouch" in return. Emphasizing semiotic exchanges, both humans and neural networks are interpreting meaning when they absorb images to create new graphics (Kowatari et al., 2009). Future research into this shared property may consider exchange structures (e.g., turn-taking and conversation dominance) and the dynamics of CMC-based human/machine collaboration (e.g., action coordination and problem-solving).

Finally, humans and machines share (as a product of the first and second properties), the potential to *influence* one another: to realize a potential to functionally matter to and manifest effects on the other. Scholarship investigating this shared property may attend to the ways that agents are capable of social and material impact. Certainly, people can influence machines, for instance through physical manipulations, design and programming, modding, and providing raw material for machine learning. Just the same, machines can influence humans. Machines induce emotional responses: humans experience increased physiological arousal when a robotic toy is being tortured (Rosenthal-von der Pütten et al., 2013) and act spitefully after feeling betrayed by a computer (Ferdig, & Mishra, 2004). How a system communicates with its users indirectly affects one's engagement with the machine, including trust and blame for (in)action (e.g., Lyons & Havig, 2014). These cognitive and affective responses by human users to machines demonstrate perceptions of accountability for machine influence. Moreover, researchers have started to debate machine's perceived worthiness of moral care and consideration—regardless of whether that is technically feasible or even desirable (Gunkel, 2017)—which indicates reciprocal effects caused by that influence.

Conclusion

Considering the evolution of operational, ontological, and phenomenological shifts in humans' and machines' roles in media communication, we advocate shifting to an agent-agnostic transmission model of communication. By endorsing a decades-old critique articulating a problematic isolating of agent types and normative roles (e.g., Giuseppe & Galiberti, 1998), we stress that humans and machines have the potential to be theoretically and operationally equivalent agents in communication processes as both social and functional actors according to the same criteria of agency, interactivity, and influence. Drawing on the HMC paradigm (principally, that meaning-making is a joint activity among human and machine agents), anthropocentric and technocentric perspectives are acknowledged as *specific* permutations of how humans and machines engage each other in the loop of transmission communication, but not the *only* permutations. Rather, humans and machines are potentially equivalent interlocutors with potentially equivalent psychological, social, and moral consequences. We see this proposal as a first step in advancing agent-agnostic perspectives in HMC. Future scholarship is encouraged to explore the ways that our propositions may (and may not) extend to more complex models of communication (e.g., transactional, interactive, and constructionist models), and investigate the dynamics and contexts by which human and machinic agents may (and may not) occupy positions in those models. Such examinations will help to ensure that communication theory keeps pace with communication technology itself lest the discipline be poorly prepared (Guzman, 2016a) and unable to deal with an ostensibly machine-inclusive future.

Author Biographies

Jaime Banks (PhD, Colorado State University) is Associate Professor in the College of Media and Communication and Texas Tech University. Her research focuses on human-technology relations with an emphasis on social robots and videogame characters, and current work focuses on the social cognition for and perception of moral agency in social robots.

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

Maartje M. A. de Graaf (PhD, University of Twente) is Assistant Professor in the Department of Information & Computing Sciences at Utrecht University. Her research addresses questions of social, emotional, and cognitive responses to robots including the societal and ethical consequences of those responses, with the end goal of influencing technology design and policy to pursue the development of socially acceptable robots that benefit society.

https://orcid.org/0000-0001-6152-552X

References

Ahern, L. (2001). Language and agency. *Annual Review of Anthropology, 30*, 109–137. https://doi.org/10.1146/annurev.anthro.30.1.109

Aldunate, N., & González-Ibáñez, R. (2017). An integrated review of emoticons in computer-mediated communication. *Frontiers in Psychology, 7,* 2061. https://doi.org/10.3389/fpsyg.2016.02061

Axelrod, R. (1973). Schema theory: An information processing model of perception and cognition. *American Political Science Review, 67*(4), 1248–1266. https://doi.org/10.2307/1956546

Banks, J. (2015). Object, Me, Symbiote, Other: A social typology of player-avatar relationships. *First Monday, 20*(2). https://doi.org/10.5210/fm.v20i2.5433

Banks, J. (2019). Theory of mind in social robots: Replication of five established human tests. *International Journal of Social Robotics* [online before print]. https://doi.org/10.1007/s12369-019-00588-x

Barfield, W., & Williams, A. (2017). Cyborgs and enhancement technology. *Philosophies, 2*(1), no. 4. https://doi.org/10.3390/philosophies2010004

Beer, D. (2009). Power through the algorithm? Participatory web cultures and the technological unconscious. *New Media & Society, 11*(6), 985–1002. https://doi.org/10.1177/1461444809336551

Bijker, W. E., Hughes, T. P., & Pinch, T. J. (1987). *The social construction of technological systems: New directions in the sociology and history of technology*. MIT Press.

Biocca, F. (1997). The cyborg's dilemma: Progressive embodiment in virtual environments. *Journal of Computer-Mediated Communication, 3*(2). https://doi.org/10.1111/j.1083-6101.1997.tb00070.x

Blomqvist, E. (2018, October 3). Five reasons why great technology is unconditionally human. *Forbes CommunityVoice*. Archived at https://web.archive.org/web/20181004123701/https://www.forbes.com/sites/forbestechcouncil/2018/10/03/five-reasons-why-great-technology-is-unconditionally-human/

Bogost, I. (2012). *Alien phenomenology, or what it's like to be a thing.* University of Minnesota Press.

Bolter, J. D. (1984). Turing's man: Western culture in the computer age. The University of North Carolina Press.

Breazeal, C. (2004). Social interactions in HRI: The robot view. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 34*(2), 181–186. https://doi.org/10.1109/tsmcc.2004.826268

Chesebro, J. W., & Bonsall, D. G. (1989). *Computer-mediated communication: Human relationships in a computerized world*. University of Alabama Press.

Cooren, F. (2004). Textual agency: How texts do things in organizational settings. *Organization, 11*, 373–393. https://doi.org/10.1177/1350508404041998

Cowan, N. (1988). Evolving conceptions of memory storage, selective attention, and their mutual constraints within the human information-processing system. *Psychological Bulletin, 104*(2), 163–191. https://doi.org/10.1037//0033-2909.104.2.163

Day, R. E. (2000). The "conduit metaphor" and the nature and politics of information studies. *Journal of the Association for Information Science and Technology, 51*(9), 805–811.

de Graaf, M. M. A. (2016). An ethical review on human-robot relationships. *International Journal of Social Robotics, 8*(4), 589–598. https://doi.org/10.1007/s12369-016-0368-5

de Graaf, M. M. A., Ben Allouch, S., & Dijk, J. A. G. M. van. (2016). Long-term evaluation of a social robot in real homes. *Interaction Studies, 17*(3), 462–491. https://doi.org/10.1075/is.17.3.08deg

De Keersmaecker, J., & Roets, A. (2017). 'Fake news': Incorrect, but hard to correct. The role of cognitive ability on the impact of false information on social impressions. *Intelligence, 65*, 107–110. https://doi.org/10.1016/j.intell.2017.10.005

Dennett, D. C. (1989). *The intentional stance.* MIT Press.

De Santis, A., Siciliano, B., De Luca, A., & Bicchi, A. (2008). An atlas of physical human-robot interaction [review]. *Mechanism and Machine Theory, 43*, 253–270. https://doi.org/10.1016/j.mechmachtheory.2007.03.003

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

Ferdig, R. E., & Mishra, P. (2004). Emotional responses to computers: Experiences in unfairness, anger, and spite. *Journal of Educational Multimedia and Hypermedia, 13*(2), 143–161.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104. https://doi.org/10.1145/2818717

Fortunati, L. (2006). User design and the democratization of the mobile phone. *First Monday*, Special Issue 7(11).

Fortunati, L. (2014). Understanding the role of mobile media in society: Models and theories. In G. Goggin & L. Hjorth (Eds.), *The Routledge companion to mobile media* (pp. 21–31). Routledge.

Fortunati, L. (2017). Robotization and the domestic sphere. *New Media & Society* [online before print]. https://doi.org/10.1177/1461444817729366

Gagliardone, I. (2016). 'Can you hear me?' Mobile–radio interactions and governance in Africa. *New Media & Society, 18*(9), 2080–2095. https://doi.org/10.1177/1461444815581148

Gergen, K. J. (1991). *The saturated self: Dilemmas of identity in contemporary life.* Basic Books.

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

Gigerenzer, G., & Goldstein, D. G. (1996). Mind as computer: Birth of a metaphor. *Creativity Research Journal, 9*(2–3), 131–144. https://doi.org/10.1080/10400419.1996.9651168

Giuseppe, R., & Galimberti, C. (1998). Computer-mediated communication: Identity and social interaction in an electronic environment. *Genetic, Social, and General Psychology Monographs, 124*(4), 434–464.

Gunkel, D. J. (2000). We are borg: Cyborgs and the subject of communication. *Communication Theory, 10*, 332–357.

Gunkel, D. J. (2012). Communication and artificial intelligence: Opportunities and challenges for the 21st century. *Communication+ 1, 1*(1).

Gunkel, D. J. (2017). The other question: Can and should robots have rights? *Ethics and Information Technology.* [online first]. https://doi.org/10.1007/s10676-017-9442-4

Guzman, A. L. (2016a). The messages of mute machines: Human-machine communication with industrial technologies. *Communication+ 1, 5*(1), 1–30.

Guzman, A. L. (2016b). Making AI safe for humans: A conversation with Siri. In R. W. Gehl, & M. Bakardjieva, (Eds.), *Socialbots and their friends* (pp. 85–101). Routledge.

Guzman, A. L. (2018). What is human-machine communication, anyway? In Guzman, A. L. (Ed.). *Human-machine communication: Rethinking communication, technology, and ourselves* (p. 1). Peter Lang.

Hakkarainen, P. (2012). 'No good for shovelling snow and carrying firewood': Social representations of computers and the internet by elderly Finnish non-users. *New Media & Society, 14*(7), 1198–1215. https://doi.org/10.1177/1461444812442663

Harada, A. (2001). Cyclodextrin-based molecular machines. *Accounts of Chemical Research, 34*(6), 456–464. https://doi.org/10.1021/ar000174l

Hasinoff, A. A. (2013). Sexting as media production: Rethinking social media and sexuality. *New Media & Society, 15*(4), 449–465. https://doi.org/10.1177/1461444812459171

Hayes, R. A., Carr, C. T., & Wohn, D. Y. (2016). One click, many meanings: Interpreting paralinguistic digital affordances in social media. *Journal of Broadcasting & Electronic Media, 60,* 171–187. https://doi.org/10.1080/08838151.2015.1127248

Heidegger, M. (1977). The question concerning technology. In *The question concerning technology and other essays* (pp. 25–26). Harper Colophon.

Hiltz, S. R., & Turroff, M. (1993). *The network nation: Human communication via computer.* The MIT Press.

Hobbs, M., Owen, S., & Gerber, L. (2017). Liquid love? Dating apps, sex, relationships and the digital transformation of intimacy. *Journal of Sociology, 53*(2), 271–284. https://doi.org/10.1177/1440783316662718

Jackson, J. (1965). Structural characteristics of norms. In I. D. Steiner & M. Fishbein (Eds.), *Current studies in social psychology* (pp. 301–309). Holt, Rinehart and Winston.

Jipson, J., & Gelman, S. A. (2007). Robots and rodents: Children's inferences about living and nonliving kinds. *Child Development, 78*, 1675–1688. https://doi.org/10.1111/j.1467-8624.2007.01095.x

Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., Ruckert, J. H., & Gill, B. (2011). The new ontological category hypothesis in human-robot interaction. *Proceedings of HRI'11* (pp. 159–160). ACM.

Kapoor, K. K., Tamilmani, K., Pana, N. P., Patil, P., Dwivedi, Y. K., & Nerur, S. (2018). Advances in social media research: Past, present and future. *Information Systems Frontiers, 20*(3), 531–558. https://doi.org/10.1007/s10796-017-9810-y

Kaptelinin, V. (2012). Activity theory. In M. Soegaard & R. F. Dam (Eds.), *Encyclopedia of human–computer interaction* (ch. 16). Aarhus: The Interaction Design Foundation.

Khurmi, R., & Gupta, J. K. (2005). *Theory of Machines* (14th ed.). Chand & Co.

Kiousis, S. (2002). Interactivity: A concept explication. *New Media & Society, 4*, 355–383. https://doi.org/10.1177/146144402320564392

Kletz, T. A. (1982). Human problems with computer control. *Process Safety Progress, 1*(4), 209–211. https://doi.org/10.1002/prsb.720010404

Kowatari, Y., Lee, S. H., Yamamura, H., Nagamori, Y., Levy, P., Yamane, S., & Yamamoto, M. (2009). Neural networks involved in artistic creativity. *Human Brain Mapping*, *30*(5), 1678–1690. https://doi.org/10.1002/hbm.20633

Krassmann, A. L., Flach, J. M., da Silva Grando, A. R. C., Tarouco, L. M., R., & Bercht, M. (2019). A process for extracting knowledge base for chatbots from text corpora. *Proceedings of IEEE Global Engineering Education Conference (EDUCON)* (pp. 322–329). https://doi.org/10.1109/EDUCON.2019.8725064

Latour, B. (1992). Where are the missing masses? The sociology of a few mundane artifacts. In W. E. Bijker & J. Law (Eds.), *Shaping technology/building society: Studies in sociotechnical change* (pp. 225–258). MIT Press.

Latour, B. (1994). Pragmatogonies: A mythical account of how humans and nonhumans swap properties. *The American Behavioral Scientist, 37*(6), 791–809. https://doi.org/10.1177/0002764294037006006

Latour, B. (2014). Agency at the time of the Anthropocene. *New Literary History, 45*(1), 1–18. https://doi.org/10.1353/nlh.2014.0003

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., Wolff, S. (1997). A brief history of the internet. *e-OTI: An International Electronic Publication of the Internet Society.* https://www.internetsociety.org/internet/history-internet/brief-history-internet

Lister, M., Dovey, J., Giddings, S., Grant, I., & Kelly, K. (2009). *New Media: A critical introduction* (2nd ed.). Routledge.

Longo, G. O. (2003). Body and technology: Continuity or discontinuity. In L. Fortunati, J. E. Katz., & R. Riccini (Eds.), *Mediating the human body: Technology, communication, and fashion* (pp. 23–30). Erlbaum.

Lyons, J. B., & Havig, P. R. (2014, June). Transparency in a human-machine context: Approaches for fostering shared awareness/intent. In *International Conference on Virtual, Augmented and Mixed Reality* (pp. 181–190). Springer, Cham.

MacKenzie, D., & Wajcman, J. (1999). Introductory essay: The social shaping of technology. In *The social shaping of technology*, 2nd ed (pp. 3–27). Open University Press.

Mardini, W., Khamayseh, Y., Yassein, M. B., & Khatatbeh, M. H. (2018). Mining Internet of Things for intelligent objects using genetic algorithm. *Computers and Electrical Engineering, 66*, 423–434. https://doi.org/10.1016/j.compeleceng.2017.10.010

Marx, K. (1887). *Capital: A critique of political economy, Vol. 1* (S. Moore & E. Aveling, Trans.). https://www.marxists.org/archive/marx/works/1867-c1/index-l.htm

McCroskey, J. M. (1992). An introduction to communication in the classroom. Burgess.

Mehrabian, A. (1972). *Nonverbal communication.* Aldine.

Morimoto, Y., Onoe, H., & Takeuchi, S. (2018). Biohybrid robot powered by an antagonistic pair of skeletal muscle tissues. *Science Robotics, 3*(18). https://doi.org/10.1126/scirobotics.aat4440

Myers, B. A. (1998). A brief history of human computer interaction technology. *ACM Interactions, 5*(2), 44–54. https://doi.org/10.1145/274430.274436

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

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

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

Neisser, U. (1966). Computers as tools and as metaphors. In C. R. Dechert (Ed.), *The social impact of cybernetics* (pp. 71–94). Simon & Schuster.

Neuman, Y. (2006). A theory of meaning. *Information Sciences, 10,* 1435–1449. https://doi.org/10.1016/j.ins.2005.03.006

Newell, A., & Simon, H. A. (1972). *Human problem solving.* Prentice-Hall.

Nowak, K. L. (2003). Sex categorization in computer mediated communication: Exploring the utopian promise. *Media Psychology, 5*(1), 83–103. https://doi.org/10.1207/S1532785XMEP0501_4

Nowak, K. L. (2004). The influence of anthropomorphism and agency on social judgment in virtual environments. *Journal of Computer-Mediated Communication, 9*(2). https://doi.org/10.1111/j.1083-6101.2004.tb00284.x

Pantic, M., Pentland, A., Nijholt, A., & Hunag, T. S. (2007). Human computing and machine understanding of human behavior: A survey. In T. S. Huang, A. Nijholt, M. Pantic, & A. Pentland (Eds.), *Artificial intelligence for human computing* (pp. 47–71). Springer-Verlag.

Papert, S. (1987). Information technology and education: Computer criticism vs. technocentric thinking. *Educational Researcher, 16*(1), 22–30. https://doi.org/10.3102/0013189x016001022

Pentland, A., & Liu, A. (1999). Modeling and prediction of human behavior. *Neural Computation, 11*, 229–242. https://doi.org/10.1162/089976699300016890

Pickering, A. (1995). *The mangle of practice: Time, agency, & science*. University of Illinois Press.

Plotnick, R. (2017). Force, flatness and touch without feeling: Thinking historically about haptics and buttons. *New Media & Society, 19*(10), 1632–1652. https://doi.org/10.1177/1461444817717510

Puntoni, S. (2018). Amazing machines and the quest for meaning in consumption. *Meaningful Consumption, 10*(2), 19–23.

Rafaeli, S., & Sudweeks, S. (1997). Networked interactivity. *Journal of Computer-Mediated Communication, 2*(4). https://doi.org/10.1111/j.1083-6101.1997.tb00201.x

Rainie, L., & Wellman, B. (2012). *Networked: The new social operating system.* The MIT Press.

Rosenthal-von der Pütten, A. M., Krämer, N. C., Hoffmann, L., Sobieraj, S., & Eimler, S. C. (2013). An experimental study on emotional reactions towards a robot. *International Journal of Social Robotics, 5*(1), 17–34. https://doi.org/10.1007/s12369-012-0173-8

Roth, D., Lugrin, J.-L., von Mammen, S., & Latoschik, M. E. (2018). Controllers and inputs: Masters of puppets. In J. Banks (Ed.), *Avatar, assembled: The social and technical anatomy of digital bodies* (pp. 281–290). Peter Lang.

Schmoldt, D. L. (1992). Bringing technology to the resource manager … and not the reverse. Proceedings of ASPRS/ACSM/RT 92 Technical Papers (pp. 62–75). https://www.fs.usda.gov/treesearch/pubs/107

Searle, J. R. (1990). Is the brain's mind a computer program? *Scientific American, 262*(1), 25–31.

Seltzer, M. (2014). *Bodies and machines.* Routledge.

Sen, A. (1979). Equality of what? [Lecture: Tanner lecture series on human values.] Archived at https://web.archive.org/web/20190218081737/https://tannerlectures.utah.edu/_documents/a-to-z/s/sen80.pdf

Shannon, C. E. (1948). A mathematical theory of communication. *The Bell System Technical Journal, 27*, 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Shannon, C. E., & Weaver, W. (1949). *The mathematical theory of communication*. University of Illinois Press.

Shawar, B. A., & Atwell, E. (2007, July). Fostering language learner autonomy through adaptive conversation tutors. Paper presented at Corpus Linguistics, Birmingham, UK.

Simon, H. A. (1990). Invariants of human behavior. *Annual Review of Psychology, 41*, 1–19. https://doi.org/10.1146/annurev.psych.41.1.1

Stich, S. P. (1981). Dennett on intentional systems. *Philosophical Topics, 12*(1), 39–62.

Subrahmanian, V. S., & Kumar, S. (2017). Predicting human behavior: The next frontiers. *Science, 355*(6324) 489. https://doi.org/10.1126/science.aam7032

Taylor, C. (1985). The concept of a person. In: *Philosophical papers* (Vol. 1, pp. 97–114). Cambridge University Press. https://doi.org/10.1017/cbo9781139173483.005

Taylor, S., & Todd, P. A. (1995). Understanding information technology usage: A test of competing models. *Information Systems Research, 6*(2), 144–176. https://doi.org/10.1287/isre.6.2.144

Tenkasi, R. V., & Boland, R. J. (1993). Locating meaning making in organizational learning: The narrative basis of cognition. *Research in Organizational Change and Development, 7*, 77–103.

Tidwell, J. (1999). Common ground: A pattern language for human-computer interface design. Archived at https://web.archive.org/web/20190928001239/http://www.mit.edu/~jtidwell/common_ground_onefile.html

Turkle, S. (2010). In good company? On the threshold of robotic companions. In Y. Wilks (Ed.), *Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 3–10). John Benjamins Publishing.

Van der Zeeuw, A., van Deursen, A. J. A. M., & Jansen, G. (2019). Inequalities in the social use of the Internet of things: A capital and skills perspective. *New Media and Society 21*, 1344–1361. https://doi.org/10.1177/1461444818821067

Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design*. Pennsylvania State University Press.

Wentzer, T. S., & Mattingly, C. (2018). Toward a new humanism: An approach from philosophical anthropology. *HAU: Journal of Ethnographic Theory, 8,* 144–157. https://doi.org/10.1086/698361

Williams R., & Edge, D. (1996). The social shaping of technology. *Research Policy, 25*(6), 865–899. https://doi.org/10.1016/0048-7333(96)00885-2

Woolgar, S. (1990). Configuring the user: The case of usability trials. *The Sociological Review, 38*(S1), 58–99. https://doi.org/10.1111/j.1467-954x.1990.tb03349.x

Young, J. E., Sung, J. Y., Voida, A., Sharlin, E., Igarashi, T., Christensen, H. I., Grinter, R. E. (2011). Evaluating human–robot interaction. *International Journal of Social Robotics, 3*, 53–67. https://doi.org/10.1007/s12369-010-0081-8

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.3

Ontological Boundaries Between Humans and Computers and the Implications for Human-Machine Communication

Andrea L. Guzman[1]

1 Department of Communication, Northern Illinois University, DeKalb, Illinois, United States of America

Abstract

In human-machine communication, people interact with a communication partner that is of a different ontological nature from themselves. This study examines how people conceptualize ontological differences between humans and computers and the implications of these differences for human-machine communication. Findings based on data from qualitative interviews with 73 U.S. adults regarding disembodied artificial intelligence (AI) technologies (voice-based AI assistants, automated-writing software) show that people differentiate between humans and computers based on origin of being, degree of autonomy, status as tool/tool-user, level of intelligence, emotional capabilities, and inherent flaws. In addition, these ontological boundaries are becoming increasingly blurred as technologies emulate more human-like qualities, such as emotion. This study also demonstrates how people's conceptualizations of the human-computer divide inform aspects of their interactions with communicative technologies.

Keywords: human-machine communication, ontology, artificial intelligence, computers, humanity, mobile voice assistants, automated journalism, human-computer interaction, human-robot interaction

Introduction

Within the past two decades, technology companies have introduced numerous forms of communicative technologies, programs, and devices that function in the role of a communicator by exchanging messages with people or by performing a communicative task on their behalf (e.g., chatbots, automated-writing software). People now are interacting with communicators that are of a different ontological nature from themselves, prompting scholars to ask how ontological divides between communicators shape people's interactions

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

with technology (e.g., A. P. Edwards, 2018). The goal of this study is to better understand aspects of the ontological dimensions of human-machine communication. To do so, the study focuses on two research questions: (1) How do people conceptualize the ontological differences between humans and computers? and (2) What are the implications of these ontological differences for the human-machine communication process?

This study contributes to the growing body of knowledge regarding the role of ontology in Human-Machine Communication (HMC), an emerging area of communication research focused on meaning-making between humans and machines (Guzman, 2018). In human communication, a person's behavior is guided by their interpretations of their communication partner (Pavitt, 2009), and a key factor informing those interpretations is the nature of both communicators. As Dautenhahn (2004) explains, when people interact with one another, they occupy the same ontological category (human), but when people interact with technology, each communicator belongs to a different category (human or machine). Researchers have found that people's conceptualizations of the nature of humans and machines matter for how they make sense of and interact with technology (e.g., A. P. Edwards, 2018; Sundar, 2008), but ongoing research is needed regarding how people conceptualize communicators occupying different ontological categories and the role of such conceptualizations in people's communicative behavior with technology (A. P. Edwards, 2018). This study continues this work by examining people's conceptualizations of the differences between humans and computers and how these differences inform people's perceptions of and interactions with communicative technologies.

The paper begins by explaining the study's focus on the nature of humans and computers before synthesizing foundational scholarship in artificial intelligence to identify existing ontological boundaries between people and computers. Next, HMC research regarding the nature of humans and machines is reviewed. This study was conducted by analyzing qualitative data regarding people's conceptualizations of the differences between humans and computers from two different projects regarding communicative technologies. The first part of the findings section presents the human-computer divides identified within the study, including new boundaries, while the second part explains the implications of these boundaries for human-machine communication. Finally, limitations and future research directions are presented.

Literature Review

The Nature of Humans and Computers

Human-Machine Communication as an area of research is so named to signal the difference in the nature of its communicators, with "machine" purposely used instead of "technology" as an acknowledgment of the long cultural, philosophical, and technological history regarding machines in relation to humans (see Guzman, 2018). In this study, the machine is conceptualized as a meta-ontological category encompassing multiple technologies. From antiquity, people have theorized what it is to be human in comparison and contrast to human-made items and nature (Boden, 2006; Riskin, 2007). As machines and society have evolved, so too have people's definitions of the boundary between themselves and machines (Turkle, 1984; Verbeek, 2005). The many permutations of the human-machine divide are too numerous and complex to examine here, and so this study focuses on the contemporary boundary between people and computers. The computer is positioned here as an ontological subcategory of the machine and encompasses the communicative technologies studied within HMC. Although many communicative technologies are accessed via hardware other than the computer, the computer is their technological progenitor.

From a historical and theoretical standpoint, the computer's introduction was a key turning point in the nature of machines (Turkle, 1984; Weizenbaum, 1976). In contrast to industrial machines that provide the "muscle" of manual labor, computers perform computations—"mental" processes once assumed to be unique to humans. Thus, computers challenge the ontological boundary between people and machines and, in doing so, the nature of each (Turkle, 1984).

Artificial intelligence research has further challenged the human-computer divide by using humans as prototypes for developing new technological capabilities (Boden, 2006; Dreyfus, 1999; Haugeland, 1985; Turkle, 1984). A key human capability being recreated within technology is the ability to communicate (Wachsmuth, 2008). The aim for imbuing technology with this ability is for interactions with devices to feel "natural" (Wachsmuth, 2008), similar to people's conversations with one another. The result has been technologies that are designed to exhibit human traits, such as gendered voices, and to emulate human communicative behaviors, such as using a person's name. The communicative technologies that are the subject of HMC research are the products of ongoing efforts to bridge the ontological gap between people and computers.

Human-Computer Divides

Ontologies are informed by people's worldviews (Chitty, 1997), and so the defining characteristics of the nature of humans and computers vary across individuals, cultures, and academic disciplines as do the boundaries between them. This review focuses on several key ontological divides between people and computers articulated regarding artificial intelligence, the study of which is focused on bridging these ontological boundaries (Franchi & Güzeldere, 2005).

Origin of being. This boundary is predicated on the difference in how humans and computers come into existence. Machines have a single point of origin: they are made by humans (Franchi & Güzeldere, 2005). People's perspectives on the genesis of humans, however, vary. A biological definition of humans situates people as part of the natural world and the product of evolution (Mazlish, 1993), while a spiritual perspective theorizes humans as divine beings created by a high power and superior to the rest of creation (Foerst, 2005) as the only beings to possess a soul (Minsky, 1986).

Autonomy. As an artificial object created by humans, the computer also is viewed as lacking another fundamental human trait—autonomy, the ability to act of one's volition (Boden, 2006; Suchman, 2011). Computers are programmed and, thus, viewed as being restricted to executing operations as directed by humans rather than having the free will of people. Relatedly, only humans are conceptualized as having a sense of self, an awareness of who they are in relation to the world around them (Haugeland, 1985; Mazlish, 1993).

Emotion. The most straightforward articulation of this divide is that humans can feel and express emotion, computers cannot. It is this difference that has been key to how people perceive the ability of computers to make judgments in contrast to humans. Weizenbaum (1976) argues that within modernity, reason has been defined by the pursuit of science detached from feeling. Emotions, within this modernist perspective, are conceptualized as the antithesis of thought and reason (Minsky, 1986). As expressed by Bolter (1984), "the computer 'thinks' by means of dispassionate, logical, calculation" (p. 75) and, thus, is conceptualized as embodying the scientific ideal (Weizenbaum, 1976). For those who subscribe to this ideal, decisions rendered by computers are superior to those by humans whose emotions cloud their judgments.

However, it should be noted that such a view of the relationship between emotions and judgment has long been contested by some AI scholars (e.g., Minsky, 1986; Weizenbaum, 1976) and within more contemporary research within the social sciences (e.g., D. Evans, 2010; Vincent & Fortunati, 2009). Emotion is a critical component of the communication process and how people interact with and make sense of their world, including the emotion people experience that is mediated by technology (Vincent & Fortunati, 2009).

Intelligence. The introduction of the computer brought into question what it meant to have a "mind" and who, or what, could possess a mind (Turkle, 1984). The result was the prediction and establishment of parameters for "machine intelligence" (e.g., Turing, 1950) followed by an ongoing debate regarding how to achieve such intelligence, if doing so were even possible (e.g., Boden, 2006; Dreyfus, 1999; Minsky, 1986). Within these debates, the nature of intelligence itself has been contested (Boden, 2006; Kasabov, 2008). Today the question of whether a mind can be recreated within a machine remains unresolved (e.g., "Can We Copy the Brain?," 2017).

Communication. As Peters (2012) explains, philosophers such as Aristotle and Descartes viewed communication as a defining human trait. Beginning in the 19th century, the process of communication came to be culturally perceived as a critical element of the human experience (Peters, 2012). Within this context, the formal study of communication emerged and came to focus primarily on people (e.g., Schramm, 1982). In contrast, artificial intelligence and related fields have theorized communication as the exchange of information between sender and receiver, which can be a human or machine (see Guzman, 2018). The ontological divide within AI, and now communication following the establishment of HMC (Guzman & Lewis, 2019), is not predicated upon whether people or computers communicate but on how they communicate. Within AI and related fields, research has focused on mitigating this divide between people and computers as communicators (e.g., National Aeronautics and Space Administration, 1976). A key approach in addressing this divide has been to replicate human traits within computers and to model human-computer interaction upon human communication (e.g., Licklider, 1960; Wachsmuth, 2008).

While the divides reviewed here are central within AI research, they alone cannot represent every facet of the human-computer divide. Ontological boundaries are shaped by the discipline in which they originate (J. H. Evans, 2016), and so, the preceding discussion based in the AI literature may not be representative of ontological perspectives in other fields. Recent research regarding cultural representations of technology also has found that elements of ontological definitions can vary between researchers and the public as well as overlap (e.g., Sarrica et al., 2019). Furthermore, ontologies can evolve, including in response to technology (Turkle, 1984).

The Ontology of Communicators: Human and Computer

When humans interact with other people, they rapidly take into account numerous aspects of their communication partner and this assessment, in turn, guides their communicative behavior (Pavitt, 2009). A key aspect of human communication that often is taken for granted is that this exchange is taking place between humans. This ontological sameness among communicators, Dautenhahn (2004) argues, allows for a series of assumptions to guide the interaction:

> When faced with a human in a department store, we might ask ourselves, "Who is this?" . . . Yet we know clearly *what* the person is—namely a member of the human species, which already allows us to make quite strong assumptions about his or her abilities, skills, and capacities. (p. 56)

But in human-machine communication, ontological assumptions may not be as straightforward. Dautenhahn (2004) continues, "In contrast, a robot or software agent in the role of a sales assistant leaves us widely in the open about its skills and capacities. Can it talk? Can it understand English? Does it know what the color blue is?" (p. 56). Interactions with machines, particularly initial interactions, require people to first determine the nature of the communicator, what it *is*, that then determines subsequent communication behavior (Dautenhahn, 2004). Exactly what people determine a digital interlocutor to be influences their assessment of it as a message source (Sundar & Nass, 2001) and their judgment of how it should be treated (A. P. Edwards, 2018).

In human-human communication, people draw from their experience of communicating with humans and their personal knowledge of what it is to be a human to make sense of their communication partner. Within human-machine communication, the onus is on the human to decipher the nature of the communicative technology without the advantage of having shared experiences (National Aeronautics and Space Administration, 1976). As has been discussed, technologies designed as communicators also merge traits of people and computers and, in some instances, even animals. The question, then, is how people assign ontological categories to devices and programs that bring together traits across ontological boundaries (A. P. Edwards, 2018).

In a study of consumer judgments of news sources, Sundar and Nass (2001) introduce the "computer heuristic" to explain people's preference for a computer to select a news story for them because they perceive the computer's selection to be more "random" and, thus, less biased than a human's selection. Sundar (2008) relabels the "computer heuristic" as the "machine heuristic" and explains that this heuristic results in "attributions of randomness, objectivity, and other mechanical characteristics" to a computer's actions (p. 83). Because these characteristics are associated with unbiased judgment, people drawing upon this heuristic are likely to consider the computer a credible source (Sundar, 2008). What is not immediately clear, however, is what other traits people may associate with computers that also may influence their judgment of technology.

A. P. Edwards (2018) observed that in interactions with a humanoid robot, some people acted toward the robot as they would another human, some people treated it similar to a pet, and others approached it as a computer. For A. P. Edwards, these differences hinted that people may be associating the robot with more familiar entities: humans, animals, and machines. Subsequent research by A. P. Edwards found that people think of animals and humans to be more similar to one another than a robot because they have a shared origin (biological or divine) separate from that of the human-built robot. These associations of ontological sameness and difference also had implications for how people judged other people's actions toward a robot.

Overall, people's interpretations of the nature of communicative technology is a crucial aspect of their judgment of and interactions with technology (Dautenhahn, 2004; A. P. Edwards, 2018; Sundar, 2008; Sundar & Nass, 2001). Continued study of these interpretations is important to provide an understanding of how people perceive technology as a communicative other (A. P. Edwards, 2018) and begins with establishing how people conceptualize the nature of computers and humans. Ontological categories are defined by their boundaries (Franchi & Güzeldere, 2005), and the study of the human-computer divide provides an understanding of the nature of each entity in addition to their differences (e.g., Turkle, 1984). AI research has identified key human-computer divides but given that ontological definitions can vary between scholars and the public (Sarrica et al., 2019), differ across academic fields (J. H. Evans, 2016), and evolve over time (Turkle, 1984), ongoing research is warranted. Therefore, the first research question focuses on how people differentiate between humans and computers. The second research question then focuses on the implications of these ontological divides for the process of human-machine communication.

Method

The data analyzed in this study was collected during two large-scale research projects regarding the ontology of disembodied artificial intelligence as communicator. The first project focused on voice-based, mobile AI assistants (e.g., Apple's Siri) (Guzman, 2015), while the second project, which began in 2018 and is ongoing, focuses on automated news-writing software (e.g., Automated Insights's Wordsmith). Both AI technologies can be considered disembodied because they are software without the type of physical form that research has shown to be key in people's interpretations of embodied communicators (e.g., Wachsmuth et al., 2008). Each project had specific research questions relative to the technology being studied, but the projects also had shared elements. Both projects focused on people's conceptualizations of a communicator that is of a different ontological nature than themselves and, therefore, examined how people conceptualize the ontological boundary between humans and computers. Bringing data from both projects into a single study enables more data points to be analyzed and for trends to be compared and contrasted regarding related technologies, adding to the robustness of the study's findings.

Both projects used the same protocol for sampling, data collection, and analysis. The method was semi-structured interviews because the goal of the research was to examine people's conceptualizations of novel technologies. Qualitative inquiry, generally, and interviews, in particular, are well-suited for research that is explanatory and exploratory and that foregrounds participants' knowledge (see Marshall & Rossman, 1995). Following an inductive approach that enables participants to articulate their understanding of a technology also helps to fill in gaps between expert and user knowledge (Lee et al., 2014).

Purposive sampling was used in both projects. The researcher recruited participants in-person from public places (e.g., libraries, transit stations) within a large city in the United States' Midwest. Participants were selected based on factors germane to each project, including technology use (i.e., people using a certain technology or none at all), news consumption (automated news project), and demographic characteristics. Data from 46 of 64 participants within the mobile AI project and 27 of 29 people in the automated news project were analyzed for this study. The number of participants whose data is used in this study is fewer because some interviews did not discuss the nature of people and computers. The 73 participants included in this study ranged in age from 18 to 76 (M = 37.7, SD = 16.7): 59% identified as female, 40% as male, and 1% as non-binary. The racial and ethnic identity of participants was White (56%), Black (16%), Hispanic (11%), Asian (10%), Middle Eastern (3%), mixed-race (3%), and other (1%). Highest level of education attained was graduate degree (16%), some graduate school (11%), bachelor's degree (40%), associate or vocational degree (8%), some college (18%), or high school diploma (7%). Participants primarily resided in urban (67%) and suburban (23%) areas. No trends were identified within the findings related to demographic characteristics, but findings should be viewed within the context of a sample of highly-educated participants living in an urban area.

Interviews were conducted in-person at the sites where participants were recruited. Participants selected an alias for the audio-recorded interviews and were assigned a new alias for this report. The interview protocol was developed around the concept of the "active interview" (Holstein & Gubrium, 1997) that emphasizes the meaning-making process between researcher and participant. The questionnaire was developed from the literature and served as a guide for the interview with the course of each interview adapted to participant responses. This study focuses on participant answers to the question: "What is the difference between humans and computers?"

Data collection and analysis were concurrent, as is standard in qualitative research (Marshall & Rossman, 1995). The endpoint of data collection was saturation—the point at which the data being collected becomes redundant and no longer contributes to new theoretical insight (Charmaz, 2014). Within the voice assistant project, saturation was based on the project's key research questions, not individual interview questions, but analysis of responses to the interview question asking people to differentiate between humans and computers showed distinct trends and redundancy in the data, indicating question-specific saturation. Aspects of the automated news project are ongoing, but the project has progressed enough to determine that saturation had been reached regarding people's conceptualizations of the differences between humans and computers.

Data analysis also was inductive. The researcher analyzed verbatim interview transcripts using MaxQDA software. To answer this study's first research question, the researcher analyzed responses regarding the difference between humans and computers. Analysis of the separate data sets began with in-vivo coding in which individual responses were coded with a descriptive phrase using a person's own words (see Saldaña, 2013). The researcher used the in-vivo codes to identify similarities within responses and then assigned new shared codes to grouped responses reflecting the same perspective. Data sets were merged, and the researcher continued this inductive coding process focused on identifying patterns within the data regarding aspects of human-computer divides. The researcher compared these emerging patterns to the AI literature to identify points of overlap and divergence before finalizing the divides discussed here. The study's second research question was answered by using participant responses as analyzed above to contextualize their opinions toward or interactions with disembodied AI programs as documented during other aspects of participant interviews.

Findings

Conceptualizing the Human-Computer Divide

Analysis resulted in several key findings that contribute to a greater understanding of the human-computer divide. Consistent with existing research (e.g., A. P. Edwards, 2018), people's conceptualizations of ontological boundaries are not uniform, but the differences articulated in this study are wider ranging. While mirroring ontological divides within the AI literature, participants identified additional boundaries based on the relationship between tools and tool-users and the flaws inherent to a particular group, most often humans. More than half of participants identified two or more differences between people and computers. Although scholars may categorize ontological divides as distinct, in the minds of some people, they are interrelated, with one divide often serving as the foundation for another. The percentage of participants identifying particular divides is included to provide a general sense of their prevalence but should be viewed within the limitations of this study's design. Interview quotes are edited for conciseness.

Origin of being. For 18% of participants, the origin of computers separated them from humans. Participants described the computer as an object that is "developed" or "programmed" by people. For example, Jeremy explains, "For the most part, I think that, of course, you know, computers are made by humans." Left unsaid in most of these responses is the origin of humans: It is enough that computers are made by people to be ontologically separate from them. Participants who do discuss people's origins reference their biological or divine nature. For example, Brad states: "So a computer, obviously, we know where a computer comes from. Humans, no idea. I mean, you can either believe god or you can trace your evidence, but nobody knows." Such answers are consistent with the AI and HMC literature (e.g., A. P. Edwards, 2018; Foerst, 2005).

The nature of the computer as constructed object also serves as the foundation for the other ways in which people differentiate between humans and computers. For Fatimah, because people build computers, they are "smarter." In contrast, Brenda thinks the artificial nature of the computer gives it an advantage over the biological limitations of people: "Humans make computers, so I kind of think, sometimes, computers may go farther than human beings because they never get tired." Conceptualizations of human and computer origins, thus, also shape people's perspectives of other traits inherent to each.

Tools and tool users. Closely related to the human-computer divide regarding origin of being is a divide focused on the nature of computers as "tools" and of people as tool users. This divide, which reflects a specific ontological relationship between people and computers, was expressed by 14% of participants. From this perspective, computers not only are made by humans but also are made to be used by humans. As Niel states: "Computers are machines developed by humans to serve humans and other functions." Niel's statement also captures another aspect of the difference between people and computers—that as machines, computers are intended to be of service to humans, a viewpoint also expressed by Joy: "I think computers are a tool to help us assemble information, disseminate information." Although the conceptualization of computers as tools and people as tool users is not identified within the AI literature discussed above, it is reflective of perspectives of the nature of people dating back more than a century (e.g., Carlyle, 1884) and of the nature of technology prominent within Western culture (e.g., Pacey, 1983).

Autonomy. The view of computer as a tool is one of many perspectives of technology that also feed into participant's conceptualizations of the differences in the levels of autonomy between computers and people. The ability to act of one's own volition is a divide between the nature of people and computers that has been discussed by AI scholars (e.g., Boden, 2006; Suchman, 2011) and, within this study, identified by 26% of participants.

Discussions of autonomy present people as possessing "free will" while the computer is restricted because it is programmed by humans, a reference back to its origins. Talula explains that ". . . humans, for the most part, we're autonomous in the way that we think. But machines are programmed to think a certain way or to arrive to a certain answer." This programmed nature results in a functional rigidity, according to participants like Dolores: "A computer has to be told every single thing to do. A computer can't think on its own like, 'Well, that didn't work, so let me try something else.' A computer has to be programmed." In contrast, people's ability to think autonomously provides them with mental flexibility not possessed by computers. Curtis further describes the difference: "I would say a computer is told what to do, and it does it. It's like a tool. I can feel one way about something, and then you could tell me, 'Hey, you should look at it this way,' and then I might." Curtis's remarks also demonstrate the interplay among ontological boundaries: Because the computer is a programmed tool it cannot act on its own.

Intelligence. AI researchers have debated whether technology can possess human intelligence for more than 70 years (e.g., "Can We Copy the Brain?," 2017; Dreyfus, 1999), and today this question remains unresolved among scholars and, based on 44% of participants, the public as well. Participants identified multiple concepts related to intelligence, such as learning, knowledge, and decision-making, as points of delineation between people and computers. Some people defined this divide in terms of people possessing intelligence while computers do not. The majority of responses, however, reflected degrees of difference (Mazlish, 1993), with both people and computers possessing a trait but to varying degrees or as expressed differently.

Participants had wide-ranging perspectives on whether people or computers had the greater ability to acquire, process, and retain information. For Carol, humans possess greater intellect: "I think our brains are smarter than their's really." When asked to explain further, Carol adds:

> We program computers. We've made them the holder, the place holder, of all of the things that we want to have at a moment's touch. But our brains have the

> capacity to be able to hang onto all kinds of infinite bits of information . . . you think about how humanity started and the whole idea of 'how do you hold onto bits of information?'

Intellect, therefore, also is tied to the origin of humans and computers from Carol's perspective: That people have stored and passed along knowledge since the start of the species and then created computers to assist them elevates human intellect over that of the computer. Carol also uses the term "bit"—the vocabulary of the computer—in stating this difference, illustrating the argument by scholars that the way people understand and talk about the nature of humans and technology remains deeply intertwined (e.g., Agre, 1997; Mazlish, 1993).

Nicki is of a different opinion than Carol: "I feel like a computer would be more—I don't want to say smarter—but maybe more advanced than the average human." Explaining why, she says, "We can just say that maybe they [computers] have a bigger brain if you look at it that way. They're more knowledgeable." Nicki also would rather receive information from a computer because while people and computers are prone to "sources of error," the computer has less error.

The boundary between humans and computers intellectually also is built around the types of knowledge and information each possess and how they make sense of and act upon it. Flynn describes this difference in terms of acquired versus accumulated knowledge:

> You have intellect. You have intelligence. You have acquired knowledge. You know, the AI is probably based on, as I know it, accumulations of knowledge. But we learn . . . And that's one of the nice things about being a person is that as you live, you learn. I'm almost 70, and I still learn.

Both humans and computers have knowledge from Flynn's perspective, but what gives humans intelligence and sets them apart from computers is that their knowledge is the result of lived experience. In contrast, computers have compiled facts and figures but without this context.

For participants similar to Eleanor, the intellectual divide is more about the different strengths and weaknesses of computers and people than which one has superior intelligence:

> Humans can think. Computers can compute. Computers can remember things. They retain information. They can find information. They do all kinds of stuff like that very quickly and very easily, and I certainly can't do that sort of thing. But a computer can't say, 'Does this make sense?'

Eleanor's answer points to a clear, but borderline tautological, divide between how people and computers deal with information, with computers excelling at some tasks and humans at others.

Emotion: Emotion also was a prominent divide, with 44% of participants saying the ability to feel and express emotion and to share in this process separates people from computers. Rachel explains that emotion is an exclusively human trait: "I mean the difference is definitely emotions . . . being able to feel compassionate for someone else, empathy, all that, sympathy." As with other divides, the line drawn over emotion can be shaped by people's views of the autonomy and origin of humans and computers. Janet explains that the computer lacks emotion because "ultimately it's already programmed by a human. It's still a machine. It's not designed as a real, thinking person would be." As an artificial object, computers cannot have emotion.

Aware of this ontological divide, developers have designed communicative technologies with emotional cues (e.g., Breazeal, 2002). Some participants who have interacted with these technologies consider emotion to be a shared trait between people and computers. During an interview regarding Siri, Rosa states ". . . it's kind of like bridging the line between technology and human emotions. I know you can tell it jokes, and it would respond to it sometimes . . ." Regarding what Rosa considers to be the difference between people and computers, she states, "Um, I don't know. Maybe it's just [the] sense where you can actually interpret the difference in sarcasm, humor. Because there's some things you can tell Siri, and it wouldn't register." For people like Rosa, technology can express emotion but not understand it.

For other participants, the fact that computers cannot understand emotion is central to the divide. Even when a communicative technology may seem to express emotion, participants like Brad argue that what is being conveyed is "perceived emotion." Andrew also explains further: ". . . you can't really input empathy into technology that's genuine. Of course, you can have sort of false and pseudo-empathy." Similar to the nature of intelligence, which has become its own ontological debate within AI (Boden, 2006), the nature of emotion also is increasingly contested.

Emotions also continue to remain central to people's differing conceptualizations of judgments rendered by people and computers. For some participants, such as Amy, computers are better able to make certain types of decisions because they do not have emotions:

> I think machines are able to make very hard decisions and are able to see everything through the constant lens of logic and rationality, whereas humans are very irrational and erratic, and there's change to them, and they're emotional and weird and kind of unpredictable, whereas machines are predictable and logical.

From Amy's perspective, emotions interfere with human judgment and emotionless technologies are better able to make certain decisions. However, other participants think the computer's lack of emotion is problematic for its judgment. Devon explains, "So it's kind of like the King Solomon thing, like split the baby in half. That's how computers would think, but we would think, 'Oh my god, that's terrible.'" Devon is referencing a story from the Judeo-Christian tradition in which Solomon determines the rightful mother of a baby using a test of empathy. For Devon, who has a background in neuroscience, emotion and empathy provide people with a more complex understanding of the world that is vital to decision-making. Amy's remarks exemplify the cultural conceptualization of emotions as detrimental to judgment, while Devon's comments echo the counter-argument. Both signal that the debate over emotion, judgment, and technology has not abated among members of the public and may even become more complex if more people interpret technology as having emotion.

Inherent flaws. Twelve percent of the responses from participants in this study offered a perspective of the difference between humans and computers that favored technology. These responses are separate from those previously discussed that weighed the advantages and disadvantages of human and computer traits tied to larger discussions of particular divides. The responses discussed here were offered as the primary reason humans and computers are different.

When asked to identify the difference between people and computers, Trisha replies, "I think that people make mistakes and do things inefficiently, and a computer will make less mistakes and do things more efficiently." Similarly, Natalie states, "I think humans are more flawed . . . computers have their technical difficulties too, but in terms of the answers they'll give you, or the stuff they're supposed to do, humans could actually screw up a lot more in terms of that." Jaheim's initial response was to state that people do not have good intentions, adding that he has "met a few bad people" but should not "really use that as a judgment." Jaheim continues: "But I do think that computers could be more honest [than humans]. I mean pretty much if you tell a computer to do something, it will do that, and it will do that to the best of its ability." In contrast, Jaheim explains, humans "cut corners" and "try to slack."

The reason why some participants focused on the shortcomings of humans is not clear, and it is more difficult to locate these answers in ontological perspectives within the literature. However, the responses discussed here place the flaws of people, at least partly, within a task- or work-based context. Critical theorists of technology have argued that within a capitalist society, the development of performance standards around the capabilities of machines, such as efficiency, would result in people being viewed as inherently less than machines (e.g., Noble, 2011), an argument extended now to the ways in which people have to routinely verify their humanity (Fortunati et al., 2019). Such a judgment is outside the scope of this study but is a potential area for continued research.

Other divides. Human-computer divides representing less than 3% of responses include possessing a soul (human); consciousness (human); creativity (human); personality (human); mortality (human); mobility (human); and expendability (computers). Many of these divides appear in the data of only one project, potentially suggesting that certain boundaries may be associated with specific communicative technologies; however, the number of people identifying these divides is so small that it is not possible to reach a determination within this study.

Implications for Human-Machine Communication

The study's second research question focused on the implications of ontological boundaries for the human-machine communication process. The findings are drawn from participants' explanations of their judgments of or behaviors with communicative technology that were consistent with their answers to the interview question regarding the difference between humans and computers. As research has shown (e.g., Sundar, 2008; Sundar & Nass, 2001), people's interpretations of the nature of a communicator play a role in their judgments of it, including within this study. Some participants' assessments of a technology's communicative abilities reflect how people differentiate between humans and computers. In addition, aspects of some people's communicative behavior with technology can be attributed to their interpretations of the human-computer divide, including whether they choose to interact with a technology and in what context.

One of the first decisions people make in human-machine communication is whether to even interact with a communicative technology. For some people, this decision is informed by their conceptualization of the boundary between humans and computers. Dolores differentiates between people and computers based on the computer's lack of autonomy that, in turn, negatively affects its decision-making. Dolores explains that "computers are not smart enough" to engage in verbal exchanges with people because they cannot process messages that deviate from their programming, and so she avoids interacting with Siri and similar technologies. Rachel also prefers communicating with people because, unlike computers, people possess and can understand emotion: "that's what I like with humans overall compared to computers is because they have feelings." And so, some people's conceptualizations of the human-computer divide play a role in determining whether they interact with a communicative technology at all.

For other people, decisions about communicating with technology hinge on the purpose of the interaction and whether a human or computer is better suited to achieving that goal. Nicki is an avid Siri user who thinks computers are "more knowledgeable" than people. When asked if she has ever turned to Siri for information instead of a human, she says, "Oh, yeah. Or even if I do [ask a person] I'll still use her [Siri] to verify." Nicki's assessment of computers as possessing more information than people informs her decisions of when to interact with Siri instead of a person. Similarly, Curtis's perception of the human-computer divide over autonomy plays a role in his decision to interact with people and computers in different contexts:

> When it comes to certain things, I'm gonna go with the computer. If it comes to how to do something, you know, how to feel, or if I'm looking at how to teach a lesson, I wanna listen to the teacher who's been doing it for 20 years over a computer program telling me, 'You should do this.' I think it's just that gray area that humans have that you can just go, 'Hmm, okay, I can take this and that.' A computer's gonna be just what it's told to give you.

Curtis weighs the intellectual autonomy of people against the programmed computer in choosing when to interact with one type of communicator instead of the other. Curtis's decision to turn to a human regarding his profession also may be indicative of how people consider not only ontological difference but also sameness in conceptualizing communicative others (e.g., A. P. Edwards, 2018).

The ontological boundary between humans and computers can also inform people's expectations of the capabilities and behaviors of technology during the human-machine communication process. For example, Tony thinks computers lack autonomy because they are programmed tools. Tony's conceptualization of the differences between humans and computers is reflected in how he contrasts the writing capabilities of human journalists with his expectations of news stories created by news-writing programs: "The difference between the two . . . is that a human will bring in their own experiences, their own writing style, their own background . . . Whereas all the computer can do is take what's been programmed for it and bring it in." Tony perceives the software's communicative abilities will be limited because its autonomy is limited.

As discussed, Rosa perceives Siri as "bridging" the human-computer divide regarding emotion and describes these interactions with Siri as "kind of weird." Rosa explains: ". . . usually when you think about your phone, you think of technology. You think robotic. You think simple logistics. And so, when you start doing things like telling it jokes, it carries more human characteristics, I guess." What makes interactions with Siri "weird" is that Rosa is not expecting emotion to be part of a communicative exchange with a technology; this weirdness is not negative for Rosa who enjoys Siri. As A. Edwards et al. (2019) found, expectancy violations can take place in human-machine communication when the nature of the communicator or its behavior do not match a person's expectations for an interaction, as is the case in Rosa's interpretations of Siri.

People's conceptualizations of the human-computer divide play an important role in human-machine communication, including informing people's actions with or judgments of communicative technologies. As exemplified in Curtis's decision-making when selecting a communication partner, multiple factors, including ontological difference and sameness, inform people's decisions and actions in human-machine communication. Ontological boundaries, therefore, are important in people's communication with technology, but they are one of many elements shaping these interactions, as is the case with most aspects of communication generally.

Conclusion

The purpose of this study was to better understand people's conceptualizations of the ontological divide between humans and computers and the implications for human-machine communication. Many of the ontological boundaries within the foundational artificial intelligence literature—origin of being, autonomy, intelligence, and emotion—remain lines of delineation between people and computers from the perspective of the public. Some of these divides, however, are no longer as clear as they once were or are becoming even more complex. Most people consider emotion to be a key boundary, but some people's interactions with communicative technologies designed to emulate human emotions, such as Apple's Siri, have caused them to reassess the degree to which emotion remains a human trait. The debate regarding intelligence also continues; although, the question is not necessarily whether technology can be intelligent. For many people, both humans and computers have attributes of intelligence: Some people think humans are "smarter" while others think computers have bigger "brains." Relatedly, some people think aspects of human nature are inherently flawed when compared to that of computers, constituting a new divide. Based on these findings, scholars should be mindful that, although many of these divides are long-standing, they are not immune to change, particularly as technology design continues to integrate more human characteristics, such as emotion. The emergence of a new divide based on human flaws also underscores that these boundaries are not purely technological, they are social, and ongoing study is warranted to better understand the social aspects shaping these divides.

This study documents not only what people think divides humans and computers but also how people think about these divides. For most people, there is no singular ontological boundary; there are multiple divides, some of which serve as the foundation for others. Continued research is needed to better understand the relationship among these divides within people's minds, such as whether particular divides routinely serve as the basis for others, and to determine if certain boundaries are more influential than others in shaping people's interpretations of technology.

As the findings demonstrate, ontological boundaries have important implications for various aspects of the human-machine communication process. People's conceptualizations of the differences between humans and computers and the nature of each help to shape their overall interpretations of technology as a communicator. These interpretations then inform people's decisions and actions within the human-machine communication process. As documented here, perceived differences in the nature of humans and computers play a role in whether people choose to communicate with a particular technology, when people choose to communicate with a technology instead of a human, and what people expect from a technology as a communication partner. Because ontologies form the foundation of who or what something is as a communicator (Dautenhahn, 2004), it is likely that people's interactions with technology are influenced by ontological boundaries in other ways not identified here, and ongoing research is warranted to understand more fully how ontological divides between communicators shape interactions across these divides. Future research also will have to clarify the role of ontological divides as one of many factors that influence people's judgments of and actions with communicative technology.

Findings should be considered within the context of the study's limitations. The findings reported here are based on a sample of highly-educated people living in urban areas within the United States, and future research should take into account other populations or examine ontological boundaries across groups. The focus of the two research projects in which data was collected, disembodied AI technologies, also may have resulted in some boundaries being more pronounced, while others may be underrepresented or overlooked. For example, people's conceptualizations of the human-computer divide regarding mobility may be different in studies regarding embodied technologies, such as robots. Communication also was not identified as an ontological boundary, most likely because the focus of the research projects was on technologies that could carry out communicative functions. Ongoing research also is needed to determine which boundaries are universal among different types of technology and which are restricted to specific devices and programs. Finally, given the breadth of people's conceptualizations of the nature of humans and computers, HMC researchers should consider studying ontological aspects within the context of their own research. Adding questions or measures related to ontology may assist scholars in better interpreting their results.

Acknowledgments

This research was supported by a Research and Artistry Facilitation Grant from Northern Illinois University and a Faculty Research Grant from the Newspaper & Online News Division of the Association for Education in Journalism and Mass Communication.

Author Biography

Andrea L. Guzman (PhD, University of Illinois at Chicago) is an assistant professor of communication at Northern Illinois University (DeKalb, IL, USA) where she focuses on Human-Machine Communication research regarding people's perceptions of artificial intelligence technologies that function as communicators, including voice assistants and automated journalism software. Guzman's work has been published in leading journals, including *New Media & Society, Digital Journalism,* and *Computers in Human Behavior.* Guzman is editor of *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* and the inaugural chair of the Human-Machine Communication Interest Group of the International Communication Association.

https://orcid.org/0000-0002-6874-9435

References

Agre, P. E. (1997). *Computation and human experience*. Cambridge University.

Boden, M. A. (2006). *Mind as machine: A history of cognitive science* (Vol. 1). Oxford University Press.

Bolter, J. D. (1984). *Turing's man: Western culture in the computer age*. University of North Carolina Press.

Breazeal, C. (2002). *Designing socialable robots*. The MIT Press.

Can We Copy the Brain? (2017, May 31). Archived at https://web.archive.org/web/20170622220350/http://spectrum.ieee.org/static/special-report-can-we-copy-the-brain

Carlyle, T. (1884). *Sartor resartus: The life and opinions of Herr Teufelsdröckh*. In *Carlyle's Complete Works* (The Vellum Ed). John Wilson & Son, University Press.

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). SAGE.

Chitty, A. (1997). First person plural ontology and praxis. *Proceedings of the Aristotelian Society*, 97, 81–96.

Dautenhahn, K. (2004). Socially intelligent agents in human primate culture. In S. Payr & R. Trappl (Eds.), *Agent culture: Human-agent interaction in a multicultural world* (pp. 45–71). https://doi.org/10.1201/b12476

Dreyfus, H. L. (1999). *What computers still can't do: A critique of artificial reason* (6th ed.). MIT Press.

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. 29–50). https://doi.org/10.3726/b14399

Evans, D. (2010). *Emotion: A very short introduction*. Oxford University Press.

Evans, J. H. (2016). *What is a human? What the answers mean for human rights*. Oxford University Press.

Foerst, A. (2005). Artificial intelligence and theology: From mythos to logos and back. In S. Franchi & G. Güzeldere (Eds.), *Mechanical bodies, computational minds: Artificial intelligence from automata to cyborgs* (pp. 489–513). MIT Press.

Fortunati, L., Manganelli, A. M., Cavallo, F., & Honsell, F. (2019). You need to show that you are not a robot. *New Media & Society*, 1–18. https://doi.org/10.1177/1461444819831971

Franchi, S., & Güzeldere, G. (Eds.). (2005). Machinations of the mind. In *Mechanical bodies, computational minds: Artificial intelligence from automata to cyborgs* (pp. 15–151). MIT Press.

Guzman, A. L. (2015). *Imagining the Voice in the Machine: The Ontology of Digital Social Agents*. University of Illinois at Chicago, Chicago, IL.

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., & Lewis, S. C. (2019). Artificial intelligence and communication: A Human–Machine Communication research agenda. *New Media & Society*. https://doi.org/10.1177/1461444819858691

Haugeland, J. (1985). *Artificial intelligence: The very idea*. The MIT Press.

Holstein, J. A., & Gubrium, J. F. (1997). Active interviewing. In D. Silverman (Ed.), *Qualitative research: Theory, method, and practice* (pp. 113–129). SAGE.

Kasabov, N. (2008). Evolving intelligence in humans and machines: Integrative evolving connectionist systems approach. *IEEE Computational Intelligence Magazine*, *3*(3), 23–37. https://doi.org/10.1109/MCI.2008.926584

Lee, H. R., Šabanovic, S., & Stolterman, E. (2014). Stay on the boundary: Artifact analysis exploring researcher and user framing of robot design. Presented at the ACM-CHI (pp. 1471–1474). ACM Press. https://doi.org/10.1145/2556288.2557395

Licklider, J. C. R. (1960/1990). Man-computer symbiosis. In R. W. Taylor (Ed.), *In memoriam: J. C. R. Licklider 1915–1990* (pp. 1–20). Systems Research Center.

Marshall, C., & Rossman, G. B. (1995). *Designing qualitative research* (2nd ed.). SAGE.

Mazlish, B. (1993). *The fourth discontinuity*. Yale University Press.

Minsky, M. L. (1986). *The society of mind*. Simon & Schuster.

National Aeronautics and Space Administration. (1976). *A forecast of space technology 1980–2000* (No. NASA SP-387). Washington, DC.

Noble, D. F. (2011). *Forces of production: A social history of industrial automation*. Transaction Publishers.

Pacey, A. (1983). *The culture of technology* (First MIT Press). MIT Press.

Pavitt, C. (2009). Impression formation. In S. W. Littlejohn & K. A. Foss (Eds.), *Encyclopedia of communication theory* (pp. 203–206). SAGE.

Peters, J. D. (2012). *Speaking into the air: A history of the idea of communication*. University of Chicago Press.

Riskin, J. (Ed.). (2007). *Genesis redux: Essays in the history and philosophy of artificial life*. University of Chicago Press.

Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). 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*. https://doi.org/10.1108/ITP-04-2018-0203

Schramm, W. (1982). *Men, women, messages, and media: Understanding human communication* (2nd ed.). Harper & Row.

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

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

Sundar, S. S., & Nass, C. (2001). Conceptualizing sources in online news. *Journal of Communication*, *51*, 52–72. https://doi.org/10.1111/j.1460-2466.2001.tb02872.x

Turing, A. M. (1950). Computing machinery and intelligence. In D. C. Ince (Ed.), *Mechanical Intelligence* (pp. 133–160). North-Holland.

Turkle, S. (1984). *The second self*. Simon & Schuster.

Verbeek, P. P. (2005). *What things do: Philosophical reflections on technology, agency, and design* (English Translation). Pennsylvania State University Press.

Vincent., J., & Fortunati, L. (2009). *Electronic emotion: The mediation of emotion via information and communication technologies*. Peter Lang.

Wachsmuth, I. (2008). 'I, Max'—Communicating with an artificial agent. In I. Wachsmuth & G. Knoblich (Eds.), *Modeling Communication with Robots and Virtual Humans* (Vol. 4930, pp. 279–295). https://doi.org/10.1007/978-3-540-79037-2_15

Wachsmuth, I., Lenzen, M., & Knoblich, G. (Eds.). (2008). *Embodied communication in humans and machines*. Oxford University Press.

Weizenbaum, J. (1976). *Computer power and human reason: From judgment to calculation*. W.H. Freeman & Co.

HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.4

Me and My Robot Smiled at One Another: The Process of Socially Enacted Communicative Affordance in Human-Machine Communication

Carmina Rodríguez-Hidalgo[1]

1 School of Communications and Journalism, Universidad Adolfo Ibáñez, Viña del Mar, Chile

Abstract

The term affordance has been inconsistently applied both in robotics and communication. While the robotics perspective is mostly object-based, the communication science view is commonly user-based. In an attempt to bring the two perspectives together, this theoretical paper argues that social robots present new social communicative affordances emerging from a two-way relational process. I first explicate conceptual approaches of affordance in robotics and communication. Second, a model of enacted communicative affordance in the context of Human-Machine Communication (HMC) is presented. Third and last, I explain how a pivotal social robot characteristic—embodiment—plays a key role in the process of social communicative affordances in HMC, which may entail behavioral, emotional, and cognitive effects. The paper ends by presenting considerations for future affordance research in HMC.

Keywords: affordance, social robot, robotics, agency, interaction, human-machine communication

Introduction

Social robots have recently emerged as a new type of media with which we can communicate (Zhao, 2006). Social robots are quickly being adopted in our homes, they perform tasks in customer service, and can assist people with health issues or disabilities. Given their rapid spreading and their quickly improving capabilities, many regard the advent of social robots as part of the fourth industrial revolution (Cross et al., 2019). In their most sophisticated form, social robots are able to recognize, talk with, and personalize their interactions to communicate with humans (Guzman, 2018). Accordingly, in the field of human-machine

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

communication (HMC) social robots are regarded as a new type of interaction partner (e.g., Edwards et al., 2019).

Despite the spreading and importance of social robots, our understanding of HMC awaits further study (Peter & Kühne, 2018). Notably, albeit social robots increasingly acquire characteristics that, in principle, allow social or communicative uses, it is not clear whether people in fact perceive and act upon social robots as social and communicative entities. That is, we do not yet know whether social robots' characteristics do result in social communicative affordances, or action possibilities for communication (Gibson, 1979). For this reason, this article's first research question relates to what is the process of social communicative affordance formation when applied to human-machine communication (HMC). This, because we can hardly identify new social communicative affordances without understanding the process affordance formation first.

Further and beyond this, it is also not clear whether the idiosyncratic characteristics of social robots, such as their body and face, promote social and communicative uses which are comparable to the types of uses that emerge in interpersonal interactions or during media exposure. Social robots may have a social or communicative function for people, but they may fulfill this function in a different way than what is known between humans face-to-face or also when humans interact with each other through established communication technologies, such as Social Networking Sites (SNSs). Thus, it is conceivable that new forms of social interactions emerge in HMC. Wondering about how these salient social robot characteristics shape the interaction and effects from it brings this study's second research question, that is, what is the role of a social robot's embodiment in the process of social communicative affordance formation.

In the context of these pressing questions, I argue that the gap between social robots characteristics and people's social and communicative actions can be bridged by integrating a refined conceptualization of affordances into HMC. Integrating these approaches is necessary, as the term affordance has been inconsistently defined in the past. For instance, in robotics affordances have been mostly conceptualized in terms of the capabilities of robots to physically interact in their environment, such as recognizing and lifting objects (Horton et al., 2012; Paauwe et al., 2015). In contrast, communication science views affordances mostly in a human-centered way, focusing on what users can obtain from using technologies such as SNSs (boyd, 2010). While I readily acknowledge these important contributions, in our view, in the context of HMC, extant conceptualizations of affordances are rooted in a view of technology as a tool or medium of communication which does not sufficiently reflect the new reality that social robots are social communicators (Zhao, 2006). To integrate these approaches, the present article first develops a conceptualization of affordances which can be reasonably applied to HMC. Second, based on this conceptualization, I explain how physical embodiment lays in our view at the heart of enacted social communicative affordances between humans and social robots, particularly because of the sequential exchange of enhanced nonverbal communicative cues, such as haptic and audiovisual signals (e.g., voice intonation, facial expressions, physical touch). The study is structured as follows. I first explicate the conceptual underpinnings of the term affordance in both robotics and communication. Second, I examine main definitions of affordance and point to some inconsistencies in the conceptualization of the term. Third, and to shed light into the process of communicative affordance formation, I present a series of steps in a model of enacted affordance, which in our view may help better illustrate our point that affordances are *enacted*, rather than only perceived, in the particular context of HMC. Fourth, I present embodiment as a crucial element in this enacted affordance process. Fifth, and to clarify, I exemplify how embodiment is crucial to allow enacted social affordances.

Grasping the Affordance Concept

Originally stemming from ecological psychology, the concept of affordance was first mentioned by Gibson:

> an affordance is what it offers the animal, what it provides or furnishes, for good or ill. The verb to afford is found in the dictionary, the noun affordance is not. I have made it up. I mean by it something that refers to both the environment and the animal in a way that no existing term does. It implies the complementarity of the animal and the environment. (Gibson, 1986, p. 127)

According to Gibson, affordances can be regarded as a particular characteristic of the environment which automatically triggers (and allows) an animal to carry out an action. For instance, a goat facing a hill would automatically perceive it as "climbable" due to its uphill and inclined morphology and would proceed to climb it. Here, the hill's morphology and shape are key to the action that is to be carried out. By extension, in the Gibsonian sense, the environment's shape and morphology both enable and determine actions. However, scholars quickly pointed out that not only the environments' characteristics, but also the capabilities of the animal influence which actions are performed. For instance, a goat is particularly skilled at climbing hills, whereas other animals, such as cows or butterflies, aren't proficient climbers, and would thus interact with the hill in a way which fits their specific abilities. In short, we can thus note that affordances are to be understood as the relation between the characteristics of a living entity and its environment.

Affordances Applied to Technology

Norman (1999) discussed affordances in terms of the action possibilities that are perceived by actors. His focus on human actors involved a move away from fully automatic responses and a stronger emphasis on interpretative and social processes and their influence on perceived action possibilities. For instance, when humans encounter a new tool, using it is likely to be the result of a cognitive evaluation of how it might be used and/or how other people use it, instead of a hard-wired automatic process. In line with this shifted focus, a newer definition of affordances appeared: "we view the affordances of an artifact as the possibilities (for both: thinking and doing) that are signified by users during their interaction with the artifact" (Vyas et al., 2006, p. 92). It is noteworthy that this newer view suggests that affordances are linked to social and subjective processes. Indeed, the social-constructionist view of affordances puts the emphasis on users and how they construct (signify) the meaning about the artifact, which is influenced by social and cultural factors. As we discuss below, this social-constructionist view is at odds with Gibson's original conceptualization of affordances. However, it is exactly these differences between the Gibsonian and the social-constructionist view which are important to develop a comprehensive conceptualization of affordances in HMC.

The Relationship Between Affordance and Perception

As we have explained, in the Gibsonian conceptualization of affordance there exists an almost automatic relationship between perception and action. This notion of "direct perception," "sense of immediacy," or "automaticity" which is crucial in the Gibsonian framework, was critiqued by subsequent scholarship (see Stoffregen, 2003). For instance, cognitive scientists argued that perception of an object and its action possibilities could by no means be an automatic process, as perception is influenced by other factors, such as the optical abilities of subjects, or the memory of past experiences with that object (Fodor & Pylyshyn, 1981; Horton et al., 2012).

This debate on the automaticity of affordances led other scholars to develop more refined definitions. Relevantly, Norman (1999) applied affordances to technological environments. He highlighted that affordances are first perceived, then enacted. For instance, when seeing a "clickable" button on a Web interface, Norman would reject to call the button itself strictly an affordance: "those displays are not affordances; they are visual feedback that advertise the affordances: they are the perceived affordances" (Norman, 2008). Therefore, if the person perceives a button as "clickable" and clicks on it and obtains a result (such as accessing another Web page), then the affordance manifests. This affordance could be, for instance, greater accessibility to more information online.

In addition, Norman (1999) posed that perceived affordances (e.g., the perceived possibilities for action) influence how people behave toward objects, especially technological ones. For instance, in the classic example of a chair, people may sit on it, lean on it, or even throw it to others depending on how people appraise its physical and social affordances (Pols, 2012). This conceptualization of perception as apart from action has found support in recent communication scholarship, which has considered "perceived affordances" (e.g., Fox & Holt, 2018) to be more precise. Further, dealing with the issue of affordance and perception, Nagy & Neff (2015) brought the concept of "imagined affordance" to highlight the importance of imagination in the affordance formation process, as "expectations for technology that are not fully realized in conscious, rational knowledge" (p. 1). These authors implied that affordances are not only perceived, they are in a large part imagined by users. This user-centered view has been common in communication approaches to affordances, as we will see below.

The Concept of Affordances in Communication

The conception of technology acting as a communication channel between two or more people played a role in the latest conceptualizations of affordances, which viewed them as emerging from the use of technology as a medium of communication. For instance, the field of human-computer interaction (HCI) identified key perceived affordances such as interactivity, which could result from either interacting with the computer or technology (e.g., clicking on a website link and gain access to new content), or emerging from communication between two humans mediated by a computer (e.g., two humans interacting with each other through e-mail). Even more recent conceptualizations have focused on the role of technology as enabling humans to connect and spread personal content in newer ways, for instance through Social Networking Sites (SNSs). From there, the field centered on how these social network platforms empower users, allowing greater connectivity and the diffusion of their own user-generated content. For instance, the notion of the scalability affordance (boyd, 2010) is that the visibility of users' content published in SNSs can easily escalate if other users reshare and spread content through various social networks. Other proposed affordances included persistence (online expressions are automatically recorded and archived) and searchability (online content can easily be accessed through search, boyd, 2010, p. 7).

In my view, a key recent communications approach on affordances proposed by Evans and colleagues (2016) emphasizes a relational view between people's actions and technology. In their definition of affordance, Evans et al. build on the relational view proposed by Faraj & Azad (2012), whom stated that:

> an affordance is a multifaceted relational structure, not just a single attribute or property or functionality of technology artifact or the actor. That is, affordance is often realized via the enactment of several mutuality relations between the technology, the artifact and the actor. (p. 254).

In my view, this definition emphasizes that affordances emerge enacted as a result of different relationships, which this article posits are a series of interrelated steps in the communicative process.

Consequently, Evans and colleagues (2016) focused on this relational stance and provided a conceptual definition of affordance as: "'the multifaceted relational structure' (Faraj & Azad, 2012, p. 254) between an object/technology and the user that enables or constraints potential behavioral outcomes in a particular context." From these definitions, it can be inferred that communication scholars mostly adhere to a conceptualization of affordances as a "multifaceted structure" and emphasize its relational character between the object, the features of the object, and the user. Moreover, Evans and colleagues proposed three criteria to assess whether a particular action would qualify as an affordance: (a) the affordance is neither the object nor the feature of the object; (b) the proposed affordance is not an outcome; and (c) the proposed affordance has variability. For (a) the authors explain that a "feature" represents a tool or attribute (p. 39) "that enables activity by part of the user" (Smock et al., 2011, p. 2323). For instance, the authors pose that a built-in camera on a phone is a feature, a tool which can be activated by the user, while the fact that one can capture photos or video, brings about the affordance of recordability. For (b), the outcomes, they argue that for example, easily finding a picture of someone we try to contact online, searching online for it for example, is an outcome of the affordance of accessibility (i.e., online content can easily be accessed, boyd, 2010). For (c) variability, Evans and colleagues propose that the affordance of visibility may vary (i.e., a photo may become more or less visible online).

Reassessing Communicative Affordances

While I see much value in Evans et al. (2016) affordance definition and assessment criteria, such an "elastic" or interrelated conceptualization of affordances, at least when applied to HMC, may not allow for an explanation of all the constitutive elements of the perceived affordance process. Moreover, I believe that this valuable cornerstone definition provided by Evans and colleagues could be specified in three main ways. First, as I have shown in the previous sections, this definition could include user perception, as it is how interaction partners perceive the technology and the possibilities for action it allows, that which brings affordances. Here, I take distance from this "direct link" theorized by Gibson between perception and action, and are lenient to more current approaches, which signal that any effort to measure affordances should focus on perceived affordances. Second, this definition does not consider the issue of user agency. Agency is the feeling of oneself being the initiator of an action, the sense of self obtained through the perceived control over the social world (Brandi et al., 2019). Following these authors, I argue that any communicative situation between social actors includes agency, as the source A (for instance the human), should be willing to initiate a series of social actions (e.g., eye contact, making a question), and then receive feedback from source B (the social robot) to that action (e.g., returning eye contact, answering the question). Following from this, I posit that it is not enough to be aware of the potential uses of technology, users need to *want* to use a technology as part of their personal agency with the objective of fulfilling a personally relevant goal.

Third and last, Evans et al.'s (2016) definition does not really specify what outcomes are. Strictly speaking, everything can be an outcome, including affordances. For instance, persistence, argued by some scholars to be an affordance, can be an outcome from the action of image data capture. Or if a user willingly decides to never erase a social media post, content persistence is then *also* an outcome of that users' decision, apart from the particular features of the technology itself (e.g., servers with enough storage capacity to store personal content online). In this respect, I would like to specify that this paper defines outcomes in the context of HMC as the behavioral, attitudinal, and cognitive effects emerging from interactions with social robots. As we have seen, the concept of affordance has faced various definitions and approaches, which are not usually consistent. One example of a different approach is that of robotics, which I proceed to briefly discuss hereunder.

The Concept of Affordance in Robotics

The traditional application of affordances in robotics has focused heavily on how the robot moves and physically deals with objects and elements in the environment. For instance, how does the robot successfully enact movement (walk, jump, run, e.g., Kuindersma et al., 2016), how it can distinguish different elements, such as objects, people, faces, landscapes (e.g., Dag et al., 2010), or how it can perform fine motor skills activities (e.g., push, lift, grasps objects) (e.g., Detry et al., 2011). Equally important for the robot to function well in the environment is the ability to avoid certain objects, which has been termed the traversability affordance (Uğur & Şahin, 2010). In other words, because sensing, planning, and executing are three major processes that robots must carry out to implement proper short-term responses and execute tasks in their environment (Brooks, 1986), the field of robotics has conceptualized affordances mostly in terms of these functions. Although recent approaches have considered how robots should deal with humans in the environment, however, this work is nonetheless still focused on physical domains (i.e., how a robot can use wheels to bypass or transit near a human, Lindler & Eschenback, 2011).

In my view, though valuable, these approaches leave aside a more comprehensive perspective which considers the robots' capacity to act socially toward other social actors in the environment. Robots have recently undergone significant developments and have acquired the capacity to socially interact with humans and provide meaningful behaviors and responses (Guzman, 2018; Zhao, 2006). In line with previous scholars who have proposed to focus on the processes and effects emerging from the interaction between humans and robots (e.g., Edwards et al., 2019), I argue here that these new capabilities bring a new social affordance to social robots, that is their capacity to sustain meaningful social interactions with humans. To make this point, I bring together affordance approaches from computer-mediated communication (CMC) and human-robot interaction (HRI), to exemplify how these approaches see technology as a medium or tool through which humans can communicate.

Bringing Approaches Together

As I have shown, the concept of affordances has faced inconsistent focus in two importantly related fields relevant to HMC, such as robotics and communication. While one is object-based, the other has been fully user-based. Further, even though we see enormous value in the relational approach to affordances such as those of Evans and colleagues (2016), I have shown how this approach could be made more specific in three important ways. Therefore, because this study's goal is to provide a more specific definition of the process of affordance formation, as it follows from the introduction to formally ask:

> *R.Q.1* = what is the process of enacted affordance formation applied to human-machine communication (HMC)?

As establishing conceptual clarity of the process of enacted communicative affordances represents a first step toward conceptualizing and identifying characteristics of social robots which may importantly influence affordances, we can deal now with the second goal of this theoretical article, which is to identify some crucial characteristics of social robots which may bring forth affordances. A characteristic which has been considered to be crucial in the definition of a social robot is that of physical embodiment. "Social robots are embodied agents . . . able to recognize each other and engage in social interactions" (Fong et al., 2003, p. 144). This body enables them to both perceive and act socially in their environment (Paauwe et al., 2015). Further, an agents' corporality is intrinsically related to how the body allows the robot not only to sense its environment and act in response to it, but also to exert an action as an agent *in* the environment (Wiltshire et al., 2013).

In addition, physical embodiment appears to be closely tied to a sense of agency, which is linked to sensorimotor processes such as touch and movement. Touch and physical proximity allow a social robot to enter the persons' intimate physical sphere (Altman & Taylor, 1973) and to be able to share interpersonal touch with humans. These haptic cues are essential to create a feeling of bonding, emotional warmth, and intimacy (Knapp & Hall, 1992). These arguments may evidence that the social affordances of HMC would follow a different process compared to, for instance, communicating with a computer. Further, several studies have found that physically embodied robots appear to be more engaging and compelling to communicate with compared to an avatar, for instance (e.g., Kiesler et al., 2008). Following these considerations, this study's second research question is formally asked, following from the introduction:

> *R.Q.2* = what is the role of the embodiment in the process of enacted social robots' affordances?

First Research Question: The Process of Enacted Communicative Affordance

This article's first research question asked about the process of affordance formation, applied to human-machine communication, that is the human communicating with a social robot and vice versa. I consider this a two-way process and therefore the social robot is considered as an interaction partner with equal social standing than the human. Building on the affordance conceptualization by Faraj & Azad (2012) and Evans et al. (2016), I provide here a definition of communicational affordances as both perceived and enacted possibilities for social interaction in a two-way iterative communication process, which emerges in the enactment of an integrated, sequential relational system which brings attitudinal, cognitive, and behavioral effects in both communication partners**.** Relevantly, because this is a communicational model in the context of HMC, I consider social affordances as emerging from a process of perceived and enacted communicative behaviors or actions between both partners, either the human or the social robot, which they perform in sequential fashion following an interaction and enabled by both their perception and their agency. In the case of the social robot, social agency is achieved through choosing from a myriad of interaction possibilities to respond or to initiate interaction with a human. Already, when facing communication with a human, social robots must choose between several alternatives which have been programmed. In the future, as social robots become more autonomous and sophisticated, their sense of all these interrelated steps will have increased importance and may become smoother and more automatized.

This paper definition of affordance considers communicational affordances as socially "enacted" because in my view, it does not suffice to just perceive possibilities for action with an object, it is actually the social actor (the person or the robot), who should ultimately enact the behavior to obtain a result from this social action. For instance, it does not suffice to perceive the social robot as a possible interaction partner, nor our intention and desire to talk to it which brings forth social affordances of meaningful social interaction. It is the combination of these factors plus their enactment which brings forth the social affordance of interactivity, for instance. For further clarity, I posit that the process outlined in Figure 1 would bring forth, "enact" a social affordance in the context of HMC.

Following the theoretical considerations presented in this paper, I first provide an account of the steps in the process of social affordance formation: (a) what the social agent perceives in the environment (the characteristics and features of the subject); (b) what that subject represents to the agent in terms of action possibilities, or the actions that the social agent perceives or imagines it can perform with the subject; (c) the actions that the social agent is willing to carry out (agency); (d) carrying out the interaction. After performing the social action, subsequently, the response of the second agent goes through the same sequence of steps (e, f, g, h). If the social agent decides to respond (h), then comes the result of the social action for the first agent (i).

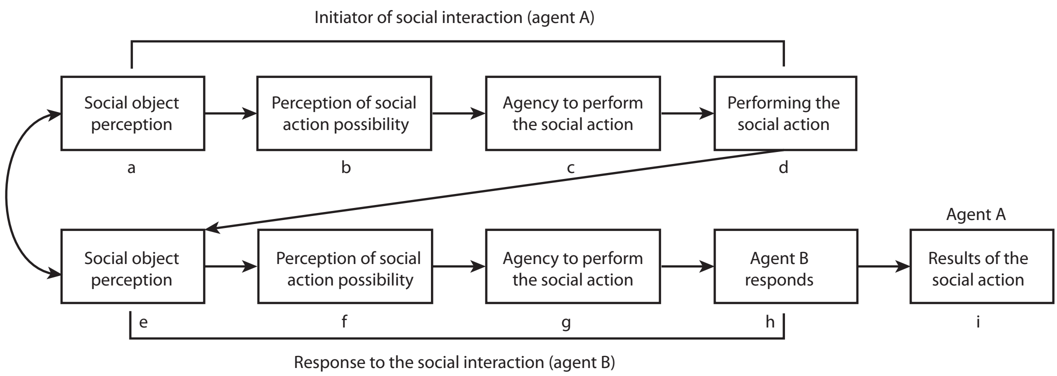


FIGURE 1 Model of Enacted Social Communicative Affordance Between a Human and a Social Robot and Vice Versa

The model of the process of enacted social affordance between a human and a social robot and vice versa, is seen on Figure 1. The semi-circular double-sided arrows to the left signal that this process is iterative, that is to say, after a social action is started by one social agent A (either the human or the social robot), the other agent B can choose to respond to the interaction and so on subsequently. First, the process takes place from the upper left square to the right. Then, as was already stated, after the social action by agent A is performed, the social agent B perceives this enacted social action and goes through the same process, from the down left square toward the right. All this process continues to the right until agent B responds. Ultimately, this response from agent B ends up by having an outcome on agent A. We exemplify the model in more detail within the next research question. I approach research question 2, regarding the role of embodiment in the process of enacted affordance.

Second Research Question: The Role of Embodiment in Social Affordances

This study's second research question asked about the role of the embodiment in the process of enacted social robots' affordance. Importantly, this paper views stimulus or robot characteristics not as affordances. Rather, these characteristics represent stimuli which influence the process of communicative enacted affordances by making some aspects more salient (e.g., presence of a physical body, presence of a mouth). Although providing a complete account of the role of embodiment is out of scope in the present theoretical study, it is used here to exemplify the model of enacted social communicative affordance below.

a) What the social agent perceives in the environment (e.g., the characteristics and features of the subject).

   Both the person and the social robot possess a body and social attributes such as body shape (e.g., presence of arms or legs), height, and facial features such as eyes and mouth. These elements will imply that both agents simply notice each other in a shared physical space.

b) What the social object represents to the agent in terms of action possibilities.

   The physical characteristics of both agents makes it that both see each other as possible social interaction partners. Considering that agent A sees agent B as a social agent and vice versa, this implies a number of perceived or imagined interaction possibilities; for instance, physically approaching the subject, waving, making eye contact, or initiating a voice-based conversation.

c) The actions that the social agent is willing to carry out (agency).

   After noticing distinct action possibilities and considering response possibilities of the other agent B, both agents will use their sense of agency to either begin an interaction or to respond to it. We posit that this agency will respond to varied personal goals of both agents, which range from socializing or learning new information. Embodiment plays a role here, because it enables action possibilities which include physicality; for instance, giving each other a hand, patting on the shoulder, or getting more proximally close.

d) Carrying out the social action.

   Agent A initiates the social action, which for example can be making eye contact, touching agent B, or initiating a dialogue. Embodiment enables these possibilities for social action because it crucially allows for embodied nonverbal language which can be encoded and decoded by both interaction partners. Crucially, it allows for physical proximity and touch of both communication partners. The robot is able to use its body to interact or send interaction signals. For instance, in the case of interacting with an anthropomorphic robot the agent has a number of "symbolic" features such as eyes or mouth to imply a certain nonverbal expression (i.e., interest), or can reach its arms to touch the person, or its legs to walk toward the person.

e) The results of this social action.

   Lastly, the social action would bring a certain outcome. Agent B will respond to the social initiation by agent A and go through the same process as agent A to initiate a response. This response will trigger an outcome in agent A. Importantly, outcomes are conceived here as the behavioral, attitudinal/emotional, and cognitive effects emerging from the social interaction. Embodiment plays a key role in the behavioral component I argue, because if for instance the conversation is running smoothly and agent A receives a suitable response from agent B, a behavioral reaction of agent A could be to decrease body distance between both agents, since they may increase their liking of the social robot. The emotional component is also affected by embodiment, as both parties presumably have a face and body, which can be used to express and read emotions by means of facial expressions in the other. Both agents' facial expressions and verbal communication have the potential of creating emotional outcomes in both partners. Lastly, cognitive effects are possible, provided that both agents, for example, cognitively process the interaction with one another, and for instance agent A, the human, learned new information thanks to the answer from agent B. An agent B can "read" and interpret both the verbal and nonverbal language provided by agent A and thus adapt their interaction depending on its goal.

Discussion

This theoretical paper had as its main aim to describe and define social communicative affordances with a focus on their formative process, in the setting of a social interaction between a human and a social robot. Further, it meant to identify a series of steps which may present themselves in the process of enacted communicative affordances in HMC. Although the present approach clearly has limitations—an important one being that its conceptualization is purely theoretical—I believe that a strength of our approach is that of presenting a comprehensive yet detailed account of the process of enacted communicative affordance formation in the context of HMC, considering the contributions of several affordance approaches. I am convinced that this approach can illuminate and enrich the discourse and research on HMC, for instance by distinguishing between perceived and enacted affordance and by presenting a scheme of the overall process of communicative affordance.

An innovative aspect of the present approach is that it considers the robot in equal social standing than the human in a communicative situation. Although this idea may seem unrealistic to some nowadays considering current social robot capabilities, particularly regarding their agency and nonverbal language expression, we believe that as social robots will become even more autonomous and social in the future, that the presented approach can be applied to better understand the new social affordances brought by social robots and can be of significant value to conceptualize perceived communicative affordances and assess effects from enacted affordances of HMC in the future.

The contributions of this study can be better understood by highlighting three conceptual remarks. These are: (1) stimulus or robot characteristics are not affordances; (2) robot characteristics influence possible enacted affordances; (3) the process of enacted affordances influence communication or interaction with robots and HMC outcomes. With regards to (1), it is not our aim to say that embodiment is an affordance per se, but rather that it is a salient social robot stimulus characteristic which is separate from perceived and enacted affordance. Although I am aware that the notion of embodiment may be rather controversial as accounting for social effects in robotics, it has been suggested that the very notion of an intelligent autonomous machine cannot exist without a body. "Intelligence requires a physical instantiation a body" (Pfeifer & Scheier, 1999). Having a body would thus integrate the conceptualizations of possessing a social intelligence, and entails occupying the same physical 3-D space with the interaction partner and further allows for visual, auditory, and especially haptic nonverbal communication, all key aspects to make social robots seem more personable and able to establish meaningful social relationships with humans.

With regards to (2), as I have postulated, this paper proposes that social robots possess unique characteristics which may influence both perceived and enacted affordances. Examples have been provided regarding embodiment, but it is my intention to leave open for further research to identify which other unique social robot characteristics render which social affordances could be enacted through HMC. As for (3), and crucial to this study's proposition, affordances influence interactions with social robots and HMC outcomes. For instance, there is ground to assume that the affordances of interactivity and customization may bring a number of behavioral outcomes, such as more frequent interaction with the social robot, emotional outcomes such as greater closeness and liking, and/or lastly, cognitive outcomes such as learning.

Future work could test and/or expand this approach by attempting to establish the empirical relationship between perceived and enacted affordances in the context of HMC. Although this paper does not propose per se that affordances are measurable (a view toward which other scholars seem skeptical, considering the concept's relational structure), this paper poses that considering the steps of the affordance formation process (e.g., making the distinction between perceived and enacted affordance and the formation steps throughout), may be meaningful and at least theoretically relevant. In a similar fashion, future scholarship could consider whether and to what extent affordances commonly discussed in the context of other media (e.g., the locatability and portability of cell phones) (Schrock, 2015), play a role in the communicative affordance formation process of social robots. Lastly, to increase our knowledge in the field, this social affordance process could be discussed and researched in the context of communicating with other nonhuman forms of communicative agents, such as algorithms or virtual AI assistants. A last and important issue to consider is the context and factors in which enacted communicative affordances and their possible effects can be explicated. Obviously, situational, environmental, individual factors, and predispositions can affect enacted affordances. Ultimately, I am hopeful that by presenting a focused yet relational perspective to enacted affordances, this may help with comprehending how people interact and what they obtain from social robots, namely the capacity of meaningful social interaction. Important here is the notion that both interaction partners ascribe meaning to socially communicating with one another (Guzman, 2018). This is important to generate behavioral, emotional, and cognitive outcomes from interacting with social robots, outcomes which are very relevant to study now and in the future.

Acknowledgments

The author would like to thank Dr. Rinaldo Kühne from the Amsterdam School of Communication Research (ASCoR), University of Amsterdam, for his insightful comments throughout various stages of manuscript development. The author declares to have no conflict of interest.

Author Biography

Carmina Rodríguez-Hidalgo, PhD, MSc, obtained her doctorate and research master degree in Communication Science at the University of Amsterdam, The Netherlands. Her research focuses on the emotional outcomes of mediated communication in youth and children. She recently joined Universidad Adolfo Ibáñez as an assistant professor in her native Chile to continue research on human-machine communication.

https://orcid.org/0000-0002-1690-9572

References

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

boyd, D. (2010). Social network sites as networked publics: Affordances, dynamics, and implications. In *A networked self* (pp. 47–66). Routledge.

Brandi, M. L., Kaifel, D., Bolis, D., & Schilbach, L. (2019). The interactive self—A review on simulating social interactions to understand the mechanisms of social agency. *i-com*, *18*(1), 17–31. https://doi.org/10.1515/icom-2018-0018

Brooks, R. (1986). A robust layered control system for a mobile robot. *IEEE journal on robotics and automation*, *2*(1), 14–23.

Cross, E. S., Hortensius, R., & Wykowska, A. (2019). From social brains to social robots: Applying neurocognitive insights to human–robot interaction. Phil. Trans. R. Soc. B 374: 20180024. https://doi.org/10.1098/rstb.2018.0024

Dag, N., Atil, I., Kalkan, S., & Sahin, E. (2010, August). Learning affordances for categorizing objects and their properties. In *2010 20th International Conference on Pattern Recognition* (pp. 3089–3092). IEEE.

Detry, R., Kraft, D., Kroemer, O., Bodenhagen, L., Peters, J., Krüger, N., & Piater, J. (2011). Learning grasp affordance densities. *Paladyn, Journal of Behavioral Robotics*, *2*(1), 1–17. https://doi.org/10.2478/s13230-011-0012-x

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. K., Pearce, K. E., Vitak, J., & Treem, J. W. (2016). Explicating affordances: A conceptual framework for understanding affordances in communication research. *Journal of Computer-Mediated Communication*, *22*(1), 35–52. https://doi.org/10.1111/jcc4.12180

Faraj, S., & Azad, B. (2012). The materiality of technology: An affordance perspective. In Leonardi, P., Nardi, B., and Kallinikos, J. (Eds.), *Materiality and organizing: Social interaction in a technological world*, Oxford University Press, pp. 237–258.

Fodor, J. A., & Pylyshyn, Z. W. (1981). How direct is visual perception? Some reflections on Gibson's "ecological approach." *Cognition*, *9*(2), 139–196.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems*, *42*(3–4), 143–166.

Fox, J., & Holt, L. F. (2018). Fear of isolation and perceived affordances: The spiral of silence on social networking sites regarding police discrimination. *Mass Communication and Society*, *21*(5), 533–554. https://doi.org/10.1080/15205436.2018.1442480

Gibson, J. J. (1986). The ecological approach to the visual perception of pictures. Original work published 1979. New Jersey: Lawrence Erlbaum Associates.

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.

Horton, T. E., Chakraborty, A., & Amant, R. S. (2012). Affordances for robots: A brief survey. *AVANT. Pismo Awangardy Filozoficzno-Naukowej*, *2*, 70–84.

Kiesler, S., Powers, A., Fussell, S. R., & Torrey, C. (2008). Anthropomorphic interactions with a robot and robot–like agent. *Social Cognition*, *26*(2), 169–181. https://doi.org/10.1521/soco.2008.26.2.169

Knapp, M. L., & Hall, J. A. (1992). Nonverbal Communication in Human Interaction. Wadsworth.

Kuindersma, S., Deits, R., Fallon, M., Valenzuela, A., Dai, H., Permenter, F., Koolen, T., Marion, P., & Tedrake, R. (2016). Optimization-based locomotion planning, estimation, and control design for the atlas humanoid robot. *Autonomous Robots*, *40*(3), 429–455. https://doi.org/10.1007/s10514-015-9479-3

Lindner, F., & Eschenbach, C. (2011, September). Towards a formalization of social spaces for socially aware robots. In *International Conference on Spatial Information Theory* (pp. 283–303). Springer, Berlin, Heidelberg.

Nagy, P., & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. *Social Media + Society*, *1*(2), https://doi.org/10.1177/2056305115603385

Norman, D. A. (1999). Affordance, conventions, and design. *interactions*, *6*(3), 38–43.

Norman, D. A. (November 17, 2008). Affordance, conventions and design. Archived at https://web.archive.org/web/20190119062638/https://jnd.org/affordance_conventions_and_design_part_2/

Paauwe, R. A., Hoorn, J. F., Konijn, E. A., & Keyson, D. V. (2015). Designing robot embodiments for social interaction: Affordances topple realism and aesthetics. *International Journal of Social Robotics*, *7*(5), 697–708. https://doi.org/10.1007/s12369-015-0301-3

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

Pfeifer, R., & Scheier, C. (1999). Understanding intelligence. MIT Press.

Pols, A. J. (2012). Characterising affordances: The descriptions-of-affordances-model. *Design Studies*, *33*(2), 113–125. https://doi.org/10.1016/j.destud.2011.07.007

Schrock, A. R. (2015). Communicative affordances of mobile media: Portability, availability, locatability, and multimediality. *International Journal of Communication*, *9*, 1229–1246.

Smock, A. D., Ellison, N. B., Lampe, C., & Wohn, D. Y. (2011). Facebook as a toolkit: A uses and gratification approach to unbundling feature use. *Computers in Human Behavior*, *27*(6), 2322–2329. https://doi.org/10.1016/j.chb.2011.07.011

Stoffregen, T. A. (2003). Affordances as properties of the animal-environment system. *Ecological psychology*, *15*(2), 115–134. https://doi.org/10.1207/S15326969ECO1502_2

Uğur, E., & Şahin, E. (2010). Traversability: A case study for learning and perceiving affordances in robots. *Adaptive Behavior*, *18*(3–4), 258–284. https://doi.org/10.1177/2F1059712310370625

Vyas, D., Chisalita, C. M., & Van Der Veer, G. C. (2006, September). Affordance in interaction. In *Proceedings of the 13th European conference on Cognitive ergonomics: Trust and control in complex socio-technical systems* (pp. 92–99). ACM.

Wiltshire, T. J., Barber, D., & Fiore, S. M. (2013, September). Towards modeling social-cognitive mechanisms in robots to facilitate human-robot teaming. In *Proceedings of the human factors and ergonomics society annual meeting* (Vol. 57, No. 1, pp. 1278–1282). SAGE Publications.

Zhao, S. (2006). Humanoid social robots as a medium of communication. *New Media & Society*, *8*(3), 401–419. https://doi.org/10.1177/1461444806061951

HUMAN-MACHINE

COMMUNICATION

HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.5

Building a Stronger CASA: Extending the Computers Are Social Actors Paradigm

Andrew Gambino[1], Jesse Fox[2], and Rabindra (Robby) A. Ratan[3]

1 The Donald P. Bellisario College of Communications, Pennsylvania State University, State College, Pennsylania, United States of America
2 School of Communication, Ohio State University, Columbus, Ohio, United States of America
3 Department of Media and Information, Michigan State University, East Lansing, Michigan, United States of America

Abstract

The *computers are social actors* framework (CASA), derived from the media equation, explains how people communicate with media and machines demonstrating social potential. Many studies have challenged CASA, yet it has not been revised. We argue that CASA needs to be expanded because people have changed, technologies have changed, and the way people interact with technologies has changed. We discuss the implications of these changes and propose an extension of CASA. Whereas CASA suggests humans mindlessly apply human-human social scripts to interactions with media agents, we argue that humans may develop and apply human-media social scripts to these interactions. Our extension explains previous dissonant findings and expands scholarship regarding human-machine communication, human-computer interaction, human-robot interaction, human-agent interaction, artificial intelligence, and computer-mediated communication.

Keywords: human-machine communication, human-computer interaction, human-robot interaction, media equation, computers are social actors (CASA)

Introduction

Theoretical development that explains and predicts human responses to social technologies has reached somewhat of a stalemate. The computers are social actors framework (CASA; Nass & Moon, 2000; Nass et al., 1994), derived from the media equation (Reeves & Nass, 1996), suggests that humans treat media and computers like real people, mindlessly applying scripts for interacting with humans to interactions with social technologies.

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

CASA is often employed to guide research situated in human-machine communication (HMC), human-computer interaction (HCI), human-robot interaction (HRI), human-agent interaction (HAI), and media effects. These subfields, and the CASA research within, have grown over the past three decades due to technological advances and wider adoption of social technologies in areas such as health care, education, and the domestic sphere (e.g., Baylor, 2011; Fortunati, 2018; Kenny et al., 2008; Takayama, 2015). As computers, machines, and media have become more complex, more variable in form, and more integrated into our lives, the theoretical limitations of CASA have become more apparent.

Because of these changes, we propose an extension of CASA that accounts for ongoing changes in people, in technologies, and in the nature of their interactions. We argue that humans have developed more specified scripts for interacting with media. As a result, when humans are mindlessly interacting with media, they do not necessarily implement social scripts associated with human-human interactions as predicted by CASA. Instead, given a deeper and broader realm of experience, humans may implement scripts they have developed for interactions specific to media entities.

To support our argument, we will investigate the existing CASA literature. First, we will explain how technologies can serve as social actors and what characteristics cue human users to their social potential. Next, we will introduce CASA and review research within the paradigm, with a particular focus on studies with theoretical implications for HMC. Next, we discuss how changes in people, changes in technologies, and changes in human relationships with technologies indicate a need to revisit CASA, particularly in light of findings that challenge CASA's claims. Based on the preceding evidence, we argue for our theoretical extension to CASA: that humans may mindlessly apply human-media scripts just as they mindlessly apply human-human scripts to social interactions with technologies. In closing, we draw attention to the implications of our proposed extension of CASA, and how it enables theory building in human-machine communication.

When Machines Act Human: Media Agents as Social Interactants

As technologies have become more interactive and replaced tasks previously performed by humans, designers have attempted to minimize the cognitive effort it takes to use them. One method is to capitalize on users' existing mental models and mimic natural forms of social interaction, creating interfaces that reflect patterns of human communication to enhance usability (Nass & Brave, 2005; Shneiderman et al., 2017). Because CASA is rooted in humans' understanding of social interaction, it is well-suited to inform research on the design and implementation of social technologies across HMC, HCI, HRI, HAI, and related fields.

To engage with CASA, it is necessary to clarify the scope of the framework and its boundary conditions. Importantly, CASA does *not* apply to every machine nor every social technology; Nass and colleagues have described two essential criteria for a technology that serve as boundary conditions for CASA's application. The first is social cues. Nass and Moon (2000) stated that "individuals must be presented with an object that has enough cues to lead the person to categorize it as worthy of social responses" (p. 83). Although this implies a boundary condition of "enough" social cues, unfortunately it is not a clearly defined one.

Given that perceptions of social potential vary from person to person and situation to situation (Waytz et al., 2010), we cannot establish objective, universal parameters of what constitutes "enough" cues. For example, shapes resembling eyes and a mouth can be sufficient to trigger a social response from a baby, but given adults' more sophisticated brains, they may not perceive the same set of shapes as indicating social potential.

The second requisite characteristic is sourcing. Nass and Steuer (1993) clarified that CASA tests "whether individuals can be induced to make attributions toward computers as if the computers were autonomous sources" (p. 511). Even in naming their framework, Nass and colleagues make a meaningful choice in declaring that "computers are social actors." This distinction is important because computers and technologies in general often serve as channels or conduits for human-human communication. The ability to enact and be perceived as a source of communication, rather than merely transmit it, indicates that a technological artifact has a degree of agency and is more than merely a channel (e.g., Sundar & Nass, 2000). Thus, for the sake of clarity and specificity, we conceptualize the types of technologies relevant to CASA as *media agents*. We define a *media agent* as any technological artifact that demonstrates sufficient social cues to indicate the potential to be a source of social interaction.

Thus, we employ the conceptualization of *media agents* to distinguish these entities from machines and technologies that are beyond the scope of CASA (see Guzman, 2018; Lewis et al., 2019, footnote 2). For example, adding a humanlike cue to a simple machine (e.g., gluing googly eyes to a stapler) does not indicate sufficient potential to be a source in social interaction. Media agents encompass a wide variety of technologies such as conversational agents, including voice assistants (e.g., Siri, Alexa), embodied conversational agents, and chatbots; virtual agents (e.g., computer-controlled video game characters); smart devices with social interfaces (e.g., a smart refrigerator), including wearables (e.g., Apple watch); and social robots (e.g., Paro, Aibo). We anticipate the number and complexity of media agents to grow as advancements are made in the technologies that power them, such as natural language processing and neural networks.

The Computers Are Social Actors Framework

The CASA framework was derived from Reeves and Nass's (1996) media equation; together, these have been referred to as the social responses to communication technologies (SRCT) approach (e.g., Sundar & Nass, 2000). In their original book *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*, Reeves and Nass (1996) argue that humans' "old brains" do not have an evolved mechanism to automatically distinguish mediated representations from their real-life counterparts. When mediated representations mimic real life, humans respond to them naturally and mindlessly (Reeves & Nass, 1996). These mindless responses extend to social interactions: When media depict social characteristics, humans treat them in a social manner rather than exerting the cognitive effort to determine how to respond (Reeves & Nass, 1996). Thus, people will assign computers personality traits, apply stereotypes and norms, and make judgments and inferences as if the computers were human, even though they understand that computers are not human (Reeves & Nass, 1996).

CASA emerged from the media equation, focusing more specifically on interactions with digital technologies. CASA focuses more narrowly on conditions wherein technologies could be perceived similarly to humans: as social actors capable of agentic communication with a human user (Nass & Moon, 2000; Nass & Steuer, 1993; Nass et al., 1994).

Social Scripts

CASA argues that people respond mindlessly to media agents and thus communicate with them similarly to how they would communicate with another human. This implies that human users have a pre-existing mental model of how they would communicate with another human in a similar situation (Nass & Moon, 2000). Mental models for interacting with others are referred to as *social scripts* (Honeycutt & Bryan, 2011; Schank & Abelson, 1977).

Over time, people encounter highly similar social situations; based on these experiences, they develop knowledge structures representing common series of sequenced behaviors. These social scripts are retained in memory and activated when relevant situations arise (Schank & Abelson, 1977). Scripts provide useful heuristics that help humans navigate rote interactions, such as ordering a beer at a local bar. Existing scripts can also be extrapolated and applied to novel yet related situations, such as ordering a Jawa Juice at the Mos Eisley Cantina on Tatooine.

The heuristic activation of social scripts underlies the CASA framework. As Nass and Moon (2000) argued:

> We can conclude that individuals are responding mindlessly to computers to the extent that they apply social scripts—scripts for human-human interaction—that are inappropriate for human-computer interaction, essentially ignoring the cues that reveal the essential asocial nature of the computer. (p. 83)

In other words, rather than questioning why a computer is demonstrating social behavior, humans follow the relevant human social script that has been mindlessly activated. CASA research has investigated how a variety of social scripts for human-human interaction are applied to human interactions with media agents.

Findings Supporting CASA

Several studies have demonstrated support for the CASA framework, including dozens of studies conducted by Nass and colleagues in the late 1990s. In these studies, the media agents were typically desktop computers that interacted using preprogrammed text-based or voice-based responses. For example, when participants were placed on the same team as a computer for a task, they rated the computer more favorably than if the computer was not labeled a teammate (Nass et al., 1996). In another study, participants encountering machines with male or female voices applied sex stereotypes to the machines, rating the female-voiced machine more expert at love and relationships and the male-voiced machine more knowledgeable about computers (Nass et al., 1997). Participants prefer computers that flatter them more than those that do not (Fogg & Nass, 1997) and evaluate a computer more positively if the computer requests the evaluation rather than providing an evaluation separately on paper (Nass et al., 1994).

As computers and media agents became more complex than text on a screen, CASA retained its predictive validity in a wide range of contexts, including anthropomorphic interfaces (E. J. Lee, 2010b), embodied agents (Hoffmann et al., 2009; K. M. Lee et al., 2006), mobile phones (Carolus et al., 2018), and voice-based navigational systems (Nass et al., 2005). For example, K. M. Lee et al. (2006) found that, similar to interpersonal interactions, a key feature of positive interactions with embodied agents is their ability to communicate through touch. Nass et al. (2005) found that matching a driver's emotion to a car's voice-based emotion led to better performance in a driving simulator. The growing variety of media agents, however, led to research that supported different explanatory mechanisms for the effects of CASA (e.g., Gong & Nass, 2007).

In recent years, the CASA framework has continued to garner support in studies examining more advanced technologies. Ho et al. (2018) found that the positive effects of interpersonal emotional disclosure applied to interactions with perceived chatbots. In HRI, studies adopting the CASA framework have demonstrated the effectiveness of human persuasive strategies (S. A. Lee & Liang, 2016, 2019) and politeness (Srinivasan & Takayama, 2016). Additionally, autonomous vehicles are perceived more positively if their voice agent's sex was stereotypically matched with its style of communication (i.e., informative male or sociable female; S. Lee et al., 2019). Results from CASA-framed studies continue to inform our understanding of communication phenomena (e.g., Hoffmann et al., 2009; Von der Pütten et al., 2010).

Although these studies have found support for CASA, several studies challenge its claims (e.g., Johnson & Gardner, 2007; M. K. Lee et al., 2012; Pfeifer & Bickmore, 2011) or suggest moderating factors (e.g., Chiasson & Gutwin, 2005; E. J. Lee, 2009; E. J. Lee, 2010a; E. J. Lee, 2010b). We consider two overarching explanations for why CASA's claims are not supported. First, these studies did not establish whether participants were interacting mindlessly. It is possible that participants conducted a sort of Turing test, determined that this social agent was not a human, but a machine, and then deliberately treated it as such. A second possibility that we will explore in more depth here is that humans have developed and mindlessly apply distinct scripts for interacting with media agents that diverge from scripts for human-human interaction.

Three Decades of Changes in People, Media Agents, and Their Interactions

In the three decades since CASA was introduced, significant changes have taken place at both the societal and technological level. Because the focus of CASA is on how humans interact with emergent technologies, changes in both humans and technologies must be accounted for by predictive theory. We argue that such changes represent a shift in the sociocultural context in which CASA is applied, that these changes are ongoing, and that they drive the need to extend the CASA paradigm.

People Have Changed: Knowledge and Experiences With Media Agents

Computational technologies have become increasingly integrated into the daily lives of people across the globe in the past 30 years. Particularly in economically advanced, industrialized nations, computers have gone from one-per-household to one-per-pocket. For example, household computer ownership in the United States increased from less than 40% in 1997 to over 80% in 2016 (Ryan & Lewis, 2017). Adult smartphone ownership in the United States grew from 35% in 2011 to 81% in 2019, with overall cell phone use at 96% in 2019 (Pew Research Center, 2019). These changes coincide with sharp increases in educational attainment (Ryan & Bauman, 2016), information technology jobs (Beckhusen, 2016), and regular computer use across job sectors (Hipple & Kosanovich, 2003). Based on these changes, we conclude that people's knowledge of and experiences with computers, and media agents by extension, have increased dramatically. Given growing exposure to and familiarity with media agents in recent years, people are likely to have changed in ways that complicate CASA.

Indeed, some studies within the CASA paradigm have supported the potential importance of individual differences such as education or experience with technology. From its earliest iterations, Nass and colleagues argued that some individual differences could be important when testing CASA, including demographics (e.g., level of education) and knowledge of technology (Nass & Steuer, 1993). Related findings suggest that CASA effects are moderated by factors such as previous computer experience (Johnson & Gardner, 2007) and that a person's expectations of media agents (e.g., social robots) vary based on their experience (Horstmann & Krämer, 2019). Experience is relevant to CASA's assumption of mindlessness given that experience might determine whether a mindless or mindful response is triggered when encountering a media agent. Relatedly, characteristics of the media agent are also likely to affect the human response.

Technologies Have Changed: Anthropomorphism

Technological advances including expanded modalities for interaction, refined graphics, and faster computing power have increased the capacity for more human-like features in media agents. Many media agents have thus become more anthropomorphic in the way they behave or how they appear. *Anthropomorphism* is the perception of human traits or qualities in an entity and indicate its potential for social interaction (Breazeal, 2003; Waytz et al., 2010). People may perceive human-like appearance, sounds, or other sensory cues in a media agent (i.e., form anthropomorphism) or human-like actions (i.e., behavioral anthropomorphism; Nowak & Fox, 2018).

Anthropomorphism is a key determinant of how media agents are evaluated (Blascovich et al., 2002; Gong & Nass, 2007; E. J. Lee, 2010a, 2010b; Rosenthal-Von der Pütten & Krämer, 2014). The study of anthropomorphism within the CASA paradigm has suggested generally positive effects (de Graaf & Allouch, 2013). For example, Gong (2008) showed anthropomorphism has a positive, linear effect on perceptions of a digital representation's competence and trustworthiness. Moreover, Gong and Nass found that participants took more time processing information before making judgments of an agent with mismatched anthropomorphic features (i.e., face and voice) than an agent exhibiting matching anthropomorphic cues. This finding suggests that higher levels of anthropomorphism facilitate faster processing and may be more likely to evoke predicted CASA effects.

Interactions Between Humans and Media Agents Have Changed

Aside from their independent factors, there are two key ways that interactions between humans and media agents have changed over time. The first is tied to affordances. *Affordances* are the inherent functional attributes of an object that indicate possible actions by a user (Gibson, 1979). More specifically, *social affordances* indicate that an object has the capacity to accommodate communication (Fox & McEwan, 2017). In terms of CASA, social affordances are relevant to understanding how humans will interpret the social potential of a media agent and if they will perceive it as a source rather than a channel.

In the past three decades, the social affordances of media agents have advanced. For example, greater memory capacity and more sophisticated artificial intelligence have facilitated increasing *personalization*, meaning that media agents can offer more tailored feedback to the user (Fox & McEwan, 2017). Relevant to CASA, one study examined ongoing interactions with a social robot and found that participants interacting with a robot exhibiting higher personalization led to greater social responses (M. K. Lee et al., 2012). Additionally, children learned more over a 2-week period from a robot that personalized its behaviors than a non-personalized robot (Baxter et al., 2017).

Further, *bandwidth*, or the array of communicative cues that are enabled, has expanded greatly from the text-based interactions of the past (Fox & McEwan, 2017). For example, modern robots can convey facial expressions, gestures, haptics, and proxemic cues. Given the broadened capacities of social affordances, media agents are providing more social cues, likely activating a broader range of social scripts among human users and demonstrating greater social potential to users. Moreover, studies have found that media agents are received more positively when they display cues that are more social (Pfeifer & Bickmore, 2011) or socially appropriate (Gratch et al., 2007).

The second key factor is time. Given the increased accessibility and adoption of computers and smartphones by organizations and individuals (Hipple & Kosanovich, 2003; Pew Research Center, 2019; Ryan & Lewis, 2017), people interact with media agents far more frequently than they did three decades ago. Additionally, it has become more common for people to have repeated, ongoing interactions with media agents, such as Amazon's Alexa, to fulfill utilitarian as well as social needs (Ammari et al., 2019; McLean & Osei-Frimpong, 2019). Improvements in social robot design also aim to facilitate long-term interactions (Leite et al., 2013) and acceptance into everyday settings (de Graaf et al., 2019).

Ongoing and long-term interactions present the opportunity for individuals to develop relationships with media agents similar to those with humans (Bickmore & Cassell, 2001; Bickmore & Picard, 2005). Longitudinal studies suggest that the relationship between a human and a media agent may change through ongoing interactions (see Leite et al., 2013, for a review). For example, Kim and Lim (2019) found that participant trust and partnership with a collaborative smartphone agent did not exist initially but were developed over 2 months of interactions. Moreover, Serholt and Barendregt (2016) found that children's social responses to robots decreased over time, suggesting that human-human interaction scripts began to fade or were applied less frequently as robots did not meet children's expectations for social interaction. Collectively, changes in the social affordances of media agents and the relationships they encourage indicate that CASA may lack explanatory power for modern users, media agents, and human-media agent interactions and relationships.

Developing Scripts for Human-Media Agent Interaction

These changes provide historical rationale to readdress CASA. Additionally, considerable evidence suggests that humans perceive media agents differently from how they perceive humans (e.g., Blascovich et al., 2002; Fox et al., 2015; Krämer et al., 2012). For example, people have distinct initial expectations for interactions with media agents (Edwards, 2018; Edwards et al., 2019; Spence et al., 2014). Different expectations and responses to humans and media agents can serve as evidence to simply refute CASA; that is, we may not treat media agents like people. Alternatively, we consider that through more social, frequent, and ongoing interactions with media agents, people may develop and apply specific scripts for interactions with media agents.

Extending CASA to incorporate scripts derived from human-media agent interaction addresses counterintuitive findings, accounts for the sociotechnological changes of the last three decades, and broadens CASA's theoretical scope. Arguably, humans need different mental models and scripts specific to media agents, or social phenomena related to the media agent, to best handle unmet expectations and maximize their own efficiency navigating novel conditions of interactions with media agents (e.g., media agents lack feelings). Results of longitudinal studies provide additional evidence that humans develop and apply scripts for interactions with media agents. Responses to social cues change upon multiple interactions with media agents, which suggests the development of scripts, and the resultant responses are systematic (Baxter et al., 2017; Bickmore & Picard, 2005; Kim & Lim, 2019; Krämer et al., 2011; M. K. Lee et al., 2012; Pfeifer & Bickmore, 2011). The systematic responses to social cues, post-change, suggest that media-derived scripts, just like human-human scripts, are applied mindlessly in interactions with media agents. Over time, we have learned to acknowledge media agents and their affordances in interactions, and we have developed more nuanced scripts for interactions with media agents. Thus, we suggest extending CASA to include scripts derived from interactions with media agents.

When CASA's assertions were being formulated, human-media agent interactions were rare and lean compared to the current landscape in which media are pervasive and rich. At that time, the presentation and experience of social affordances were more limited because of the technology powering media agents. Advances in technologies such as natural language processing, neural-networks, and raw computing power allow for social affordances of modern media agents to manifest in a wider variety of forms and aptitudes. For example, unique data collection and processing power allows media agents to personalize at a qualitatively different level than humans. Hence, the study of social responses should not be restricted by a focus on human correlates or similarities. In this way, researchers can also avoid reifying face-to-face communication as the gold standard for HMC and being constrained by the limitations of human interactions (e.g., Fortunati, 2005; Spence, 2019). Instead, researchers can explore why communication with a media agent may be preferred over communication with a human. Removing this anthropocentric bias from CASA should allow researchers a means to avert their own human-centric biases and avoid the pitfall highlighted by Groom and Nass (2007): “While trying to make robots human, researchers have sometimes overlooked what makes robots special” (p. 494).

Conclusion and Future Directions

Here, we have argued to extend CASA to include scripts developed through interactions with media agents. We additionally proposed that the development of a media agent-derived script is based on the social affordances of the media agent and temporal factors of relationships with media agents. The proposed extension broadens CASA's theoretical scope and reconciles CASA with findings that suggest people do not necessarily follow human-human social scripts when interacting with media agents.

Our extension accomplishes four goals. First, it reconciles the CASA perspective with trends in digital media use and accounts for some divergent research findings wherein human-media agent interactions are not consistent with human-human scripts (e.g., Edwards et al., 2019; Srinivasan & Takayama, 2016; Takayama, 2015). Second, it increases the theoretical specificity of CASA by enabling it to account for another type of mindless interaction with media agents. Third, it mitigates some of CASA's anthropocentric bias. The fourth and crucial goal accomplished in extending CASA is promoting the examination of time as a variable in understanding the development and application of media-agent scripts.

A first step is to begin clarifying what type of scripts exist for media-agent interactions. Methodologically, identifying such scripts is likely to require more inductive approaches. As Guzman (2018) suggests, a diverse range of methods and methodologies is necessary to advance knowledge pertaining to human relationships with machines. Although CASA's experimental paradigm is useful for controlled comparisons, it is ill-suited for the exploratory research that may be required to elaborate human-media agent social scripts due to the emergent nature of scripts as well as their predicted nuance. To capture this nuance, inductive approaches may be adopted alongside traditional CASA experiments. For example, Edwards et al. (2019) analyzed open-ended responses in their experimental study. These qualitative responses suggested that interactions with a social robot elicited positive impressions through feelings of connectedness, while interactions with a human partner, following the same interaction script, felt impersonal and disconnected. Finally, given that time is a key factor in the development of scripts, we advocate for the use of longitudinal methods to test CASA's claims as well as our proposed extension.

As we develop more refined scripts through longer, more complex, and more variant interactions with media agents, these scripts may influence our interpersonal relationships. In the same way that human scripts are mindlessly applied to guide our interactions with media, over time, media scripts may be developed and mindlessly applied to our interactions with humans. A thorough and proper understanding of human communication processes and relationships may be informed by understanding how we interact with media agents. Through extension, inquiry within CASA's framework may suggest the reverse of its core prediction. Rather than treating computers like people, we may treat people like computers.

In summary, CASA has been a productive framework for studying human-machine communication, human-computer interaction, human-robot interaction, and human-agent interaction. The extended CASA we proposed reconciles counterintuitive findings, acknowledges changes over the last three decades, encourages research from more diverse methodological approaches, and does not invalidate research findings within the CASA framework. HMC research guided by this extended CASA can inform a more robust understanding of humans, machines, communication, and the human-machine relationship.

Acknowledgments

No external funding supported this research. The authors would like to express their gratitude for the service of the reviewers and editorial team, whose suggestions markedly improved the paper. Additionally, the authors would like to thank Denise Solomon, Mary Beth Oliver, and Brandon Van Der Heide for their input and feedback at different stages of development.

Author Biographies

Andrew Gambino is a University Fellow and PhD Candidate in The Donald P. Bellisario College of Communications at Pennsylvania State University. He studies communication and relationships.

https://orcid.org/0000-0001-8657-6788

Jesse Fox (PhD, Stanford University) is an Associate Professor in the School of Communication at Ohio State University. She researches all the things.

https://orcid.org/0000-0002-5018-3250

Rabindra (Robby) A. Ratan (PhD, USC Annenberg) is an associate professor and AT&T scholar in the Department of Media and Information at Michigan State University. Dr. Ratan conducts research on the effects of human-technology interaction, examining how media technologies (e.g., avatars, agents, automobiles) influence meaningful outcomes (e.g., persuasion, education, health/safety).

https://orcid.org/0000-0001-7611-8046

References

Ammari, T., Kaye, J., Tsai, J. Y., & Bentley, F. (2019). Music, search, and IoT: How people (really) use voice assistants. *ACM Transactions on Computer-Human Interaction, 26*(3), article 17. https://doi.org/10.1145/3311956

Baxter, P., Ashurst, E., Read, R., Kennedy, J., & Belpaeme, T. (2017). Robot education peers in a situated primary school study: Personalisation promotes child learning. *PloS One*, *12*, e0178126. https://doi.org/10.1371/journal.pone.0178126

Baylor, A. L. (2011). The design of motivational agents and avatars. *Educational Technology Research & Development, 59*, 291–300. https://doi.org/10.1007/s11423-011-9196-3

Beckhusen, J. (2016). *Occupations in information technology*. U.S. Census Bureau.

Bickmore, T., & Cassell, J. (2001, March). Relational agents: A model and implementation of building user trust. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 396–403). ACM. https://doi.org/10.1145/365024.365304

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction*, *12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry, 13*, 103–124. https://doi.org/10.1207/S15327965PLI1302_01

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies*, *59*, 119–155. https://doi.org/10.1016/S1071-5819(03)00018-1

Carolus, A., Schmidt, C., Schneider, F., Mayr, J., & Muench, R. (2018). Are people polite to smartphones? *Lecture Notes in Computer Science, 10902*, 500–511. https://doi.org/10.1007/978-3-319-91244-8_39

Chiasson, S., & Gutwin, C. (2005). Testing the media equation with children. In *Proceedings of SIGCHI '05 Human Factors in Computing Systems* (pp. 829–838). ACM. https://doi.org/10.1145/1054972.1055089

de Graaf, M. M., & Allouch, S. B. (2013). Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, *61*, 1476–1486. https://doi.org/10.1016/j.robot.2013.07.007

de Graaf, M. M., Allouch, S. B., & Van Dijk, J. A. (2019). Why would I use this in my home? A model of domestic social robot acceptance. *Human–Computer Interaction, 34*, 115–173. https://doi.org/10.1080/07370024.2017.1312406

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. 29–50). Peter Lang.

Edwards, A., Edwards, C., & Gambino, A. (2019). The social pragmatics of communication with social robots: Effects of robot message design logic in a regulative context. *International Journal of Social Robotics*. https://doi.org/10.1007/s12369-019-00538-7

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

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

Fortunati, L. (2005). Is body-to-body communication still the prototype? *The Information Society*, *21*, 53–61. https://doi.org/10.1080/01972240590895919

Fortunati, L. (2018). Robotization and the domestic sphere. *New Media & Society*, *20*, 2673–2690. https://doi.org/10.1177/1461444817729366

Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., & Bailenson, J. N. (2015). Avatars versus agents: A meta-analysis quantifying the effect of agency on social influence. *Human–Computer Interaction*, *30*, 401–432. https://doi.org/10.1080/07370024.2014.921494

Fox, J., & McEwan, B. (2017). Distinguishing technologies for social interaction: The Perceived Social Affordances of Communication Channels Scale. *Communication Monographs, 84,* 298–318. https://doi.org/10.1080/03637751.2017.1332418

Gibson, J. J. (1979). *The ecological approach to visual perception*. Houghlin Mifflin.

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*, 1494–1509. https://doi.org/10.1016/j.chb.2007.05.007

Gong, L., & Nass, C. (2007). When a talking-face computer agent is half-human and half-humanoid: Human identity and consistency preference. *Human Communication Research, 33,* 163–193. https://doi.org/10.1111/j.1468-2958.2007.00295.x

Gratch, J., Wang, N., Gerten, J., Fast, E., & Duffy, R. (2007). Creating rapport with virtual agents. *Lecture Notes in Computer Science, 4722*, 125–138. https://doi.org/10.1007/978-3-540-74997-4_12

Groom, V., & Nass, C. (2007). Can robots be teammates? Benchmarks in human–robot teams. *Interaction Studies*, *8*, 483–500. https://doi.org/10.1075/is.8.3.10gro

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

Hipple, S., & Kosanovich, K. (2003). Computer and Internet use at work in 2001. *Monthly Labor Review, 126*, 26–35.

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, *68*, 712–733. https://doi.org/10.1093/joc/jqy026

Hoffmann, L., Krämer, N. C., Lam-Chi, A., & Kopp, S. (2009, September). Media equation revisited: Do users show polite reactions towards an embodied agent? *Lecture Notes in Computer Science, 5773*, 159–165. https://doi.org/10.1007/978-3-642-04380-2_19

Honeycutt, J. M., & Bryan, S. P. (2011). *Scripts and communication for relationships*. Peter Lang.

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*, article 939. https://doi.org/10.3389/fpsyg.2019.00939

Johnson, D., & Gardner, J. (2007). The media equation and team formation: Further evidence for experience as a moderator. *International Journal of Human-Computer Studies*, *65*, 111–124. https://doi.org/10.1016/j.ijhcs.2006.08.007

Kenny, P., Parsons, T., Gratch, J., & Rizzo, A. (2008, July). Virtual humans for assisted health care. In *Proceedings of the 1st International Conference on Pervasive Technologies Related to Assistive Environments* (article 6). ACM.

Kim, D. J., & Lim, Y. K. (2019). Co-performing agent: Design for building user-agent partnership in learning and adaptive services. In *Proceedings of SIGCHI 2019 Human Factors in Computing Systems*. https://doi.org/10.1145/3290605.3300714

Krämer, N. C., Eimler, S., von der Pütten, A., & Payr, S. (2011). Theory of companions: What can theoretical models contribute to applications and understanding of human-robot interaction? *Applied Artificial Intelligence*, *25*, 474–502. https://doi.org/10.1080/08839514.2011.587153

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.

Lee, E. J. (2009). I like you, but I won't listen to you: Effects of rationality on affective and behavioral responses to computers that flatter. *International Journal of Human-Computer Studies*, *67*, 628–638. https://doi.org/10.1016/j.ijhcs.2009.03.003

Lee, E. J. (2010a). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior*, *26*, 665–672. https://doi.org/10.1016/j.chb.2010.01.003

Lee, E. J. (2010b). What triggers social responses to flattering computers? Experimental tests of anthropomorphism and mindlessness explanations. *Communication Research, 37,* 191–214. https://doi.org/10.1111/j.1460-2466.2008.00386.x

Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (2006). Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human–robot interaction. *International Journal of Human-Computer Studies*, *64*, 962–973. https://doi.org/10.1016/j.ijhcs.2006.05.002

Lee, M. K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., & Savetsila, S. (2012, March). Personalization in HRI: A longitudinal field experiment. In *Proceedings of HRI 2012 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 319–326). IEEE.

Lee, S., Ratan, R., & Park, T. (2019). The voice makes the car: Enhancing autonomous vehicle perceptions and adoption intention through voice agent gender and style. *Multimodal Technologies and Interaction*, *3*, article 20. https://doi.org/10.3390/mti30100200

Lee, S. A., & Liang, Y. (2016). The role of reciprocity in verbally persuasive robots. *Cyberpsychology, Behavior, and Social Networking*, *19*, 524–527. https://doi.org/10.1089/cyber.2016.0124

Lee, S. A., & Liang, Y. (2019). Robotic foot-in-the-door: Using sequential-request persuasive strategies in human-robot interaction. *Computers in Human Behavior*, *90*, 351–356. https://doi.org/10.1016/j.chb.2018.08.026

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

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

McLean, G., & Osei-Frimpong, K. (2019). Hey Alexa … examine the variables influencing the use of artificial intelligent in-home voice assistants. *Computers in Human Behavior*, *99*, 28–37. https://doi.org/10.1016/j.chb.2019.05.009

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,* 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., Jonsson, I. M., Harris, H., Reeves, B., Endo, J., Brave, S., & Takayama, L. (2005, April). Improving automotive safety by pairing driver emotion and car voice emotion. In *CHI'05 Extended Abstracts on Human Factors in Computing Systems* (pp. 1973–1976). ACM.

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

Nass, C., Moon, Y., & Green, N. (1997). Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology, 27,* 864–876.

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

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of SIGCHI '94 Human Factors in Computing Systems* (pp. 72–78). ACM. https://doi.org/10.1145/259963.260288

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

Pew Research Center. (2019). *Mobile technology and home broadband 2019*. Pew.

Pfeifer, L. M., & Bickmore, T. (2011, May). Is the media equation a flash in the pan? The durability and longevity of social responses to computers. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 777–780). ACM.

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

Rosenthal-Von Der Pütten, A. M., & Krämer, N. C. (2014). How design characteristics of robots determine evaluation and uncanny valley related responses. *Computers in Human Behavior*, *36*, 422–439. https://doi.org/10.1016/j.chb.2014.03.066

Ryan, C. L., & Bauman, K. (2016). *Educational attainment in the United States: 2016*. U.S. Census Bureau.

Ryan, C. L., & Lewis, J. M. (2017). *Computer and internet use in the United States: 2016*. U.S. Census Bureau.

Schank, R. C., & Abelson, R. P. (1977). *Scripts, plans, goals, and understanding: An inquiry into human knowledge structures*. Erlbaum.

Serholt, S., & Barendregt, W. (2016, October). Robots tutoring children: Longitudinal evaluation of social engagement in child-robot interaction. In *Proceedings of the 9th Nordic Conference on Human-Computer Interaction* (article 64). ACM. https://doi.org/10.1145/2971485.2971536

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.

Spence, P. R. (2019). Searching for questions, original thoughts, or advancing theory: Human machine communication. *Computers in Human Behavior*, *90*, 285–287. https://doi.org/10.1016/j.chb.2018.09.014

Spence, P. R., Westerman, D., Edwards, C., & Edwards, A. (2014). Welcoming our robot overlords: Initial expectations about interaction with a robot. *Communication Research Reports*, *31*, 272–280. https://doi.org/10.1080/08824096.2014.924337

Srinivasan, V., & Takayama, L. (2016). Help me please: Robot politeness strategies for soliciting help from humans. In *Proceedings of the SIGCHI '16 Human Factors in Computing Systems* (pp. 4945–4955). ACM. https://doi.org/10.1145/2858036.2858217

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

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

Von der Pütten, A. M., Krämer, N. C., Gratch, J., & Kang, S. H. (2010). It doesn't matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*, 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, *5*, 219–232. https://doi.org/10.1177/1745691610369336

HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.6

The Robot Privacy Paradox: Understanding How Privacy Concerns Shape Intentions to Use Social Robots

Christoph Lutz[1] and Aurelia Tamó-Larrieux[2]

1 Department of Communication and Culture, BI Norwegian Business School, Nordic Centre for Internet and Society, Oslo, Norway
2 Center for Information Technology, Society, and Law, University of Zurich, Zurich, Switzerland

Abstract

Conceptual research on robots and privacy has increased but we lack empirical evidence about the prevalence, antecedents, and outcomes of different privacy concerns about social robots. To fill this gap, we present a survey, testing a variety of antecedents from trust, technology adoption, and robotics scholarship. Respondents are most concerned about data protection on the manufacturer side, followed by social privacy concerns and physical concerns. Using structural equation modeling, we find a privacy paradox, where the perceived benefits of social robots override privacy concerns.

Keywords: social robots, privacy, trust, survey

Introduction

Does the privacy paradox translate to the use of social robots? In other words, is there a robot privacy paradox? In this article, we empirically investigate the link between privacy concerns and the intention to use social robots. As social robots are increasingly interacting with us in our daily environment (Fong et al., 2003; Gupta, 2015; Van den Berg, 2016), the advantages and concerns of close human-machine interaction have become a topic of public debate. A key concern triggered by human interaction with social robots is related to users' privacy (Lutz & Tamò, 2018). As social robots function based on data analysis and have greater mobility and autonomy than static devices, it is no surprise that literature has started to investigate their privacy implications on a descriptive level (Calo, 2012; Kaminski, 2015; Kaminski et al., 2017; Lutz & Tamò, 2015; Rueben, Grimm, et al., 2017; Sedenberg et al., 2016). Yet, empirical evidence on privacy concerns and privacy implications of social robots among non-experts (i.e., individuals largely unfamiliar with robots) is scarce. While

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

a few surveys have looked at trust in social robots (Alaiad & Zhou, 2014) and general attitudes toward them (Eurobarometer, 2012; Liang & Lee, 2017), privacy has mostly been discussed in conceptual terms (Calo, 2012, 2016; Lutz & Tamò, 2018; Rueben, Grimm, et al., 2017).

In this article, we present the results of a survey that aimed at understanding the general public's privacy concerns about social robots and how these concerns affect use intention. The findings point to the need to differentiate privacy types and to the important role of the social environment in shaping users' attitudes about this new technology.

In the course of the article, we first define the term "privacy" and provide an overview of previous literature on the topic of privacy in the context of social robots. We will ground the definition in previous research about online privacy and privacy in general to reach a more holistic understanding of the phenomenon. Subsequently, we will present the research model to be tested in the survey and develop the hypotheses. We then describe the survey methodology, including the sample, data analysis approach, and measurement, and present the survey results. Finally, we discuss the limitations of our approach and contextualize the findings.

Literature Review

Defining Privacy Concerns in the Context of Social Robots

Despite new technological developments and a recent surge of interest, privacy scholarship can draw on a long academic tradition (Altman, 1975; Warren & Brandeis, 1890; Westin, 1967). Today, privacy is a multidisciplinary research field. Disciplines involved in its study include communication, computer science, psychology, sociology, information systems, economy, and law (Pavlou, 2011). However, this multitude of perspectives complicates a common understanding of the central construct. As Solove (2008, p. 1) points out:

> privacy is a sweeping concept, encompassing (among other things) freedom of thought, control over one's body, solitude in one's home, control over personal information, freedom from surveillance, protection of one's reputation, and protection from searches and interrogations. Philosophers, legal theorists, and jurists have frequently lamented the great difficulty in reaching a satisfying conception of privacy.

In this article, we rely on Bygrave's (2002) distinction of privacy aspects or types and apply them in the context of social robots. The first type we consider is *physical privacy* and the notion of non-interference (dating back to the early understanding of privacy according to Warren & Brandeis, 1890). Physical privacy considerations revolve around "physical access to an individual and/or the individual's surroundings and private space" (Smith et al., 2011, p. 990). According to Calo (2012), who offers a useful privacy typology for social robots, issues related to physical privacy are linked to a robot's ability to enter physical personal spaces, such as bedrooms and bathrooms. Since robots are increasingly employed in homes, for example as household assistants, they might be exposed to sensitive or compromising situations. Similarly, robots might have access to vulnerable population groups and their habits, such as children, the elderly, and the infirm. In this sense, physical privacy in the context of social robots relates to the notion of "freedom from" (Koops et al., 2016), which incorporates the idea that an individual remains unobserved in private spaces.

Physical privacy can be distinguished from informational privacy (Smith et al., 2011). The latter follows Westin's (1967) famous definition of privacy as a means to achieve self-realization and thus being able to control information about ourselves. Informational privacy can further be divided into two subcategories: one that relates to institutional threats and one that relates to social threats (Raynes-Goldie, 2010; Young & Quan-Haase, 2013). While *institutional informational privacy* includes privacy considerations about the processing of data by institutions (e.g., robot manufacturers, government agencies, and third parties such as data brokers or cloud providers), social informational privacy revolves around the processing of data by private individuals (e.g., familiar users, hackers). Surveillance is a dominant concern with regard to informational privacy, both institutional and social (Calo, 2012). As modern robots are equipped with innovative sensors and processors, enabling more advanced observation capabilities than humans, they potentially could be used for spying and sophisticated "background" data collection (i.e., without awareness or consent by users and bystanders).

Privacy can also be understood as "a selective control of access to the self or to one's group" (Altman, 1975, p. 18). We call this group *social informational privacy*. While in this article we understand access to the self broadly, we are interested in social "freedom from" forms of privacy (Koops et al., 2016) in the sense of informational boundary management. These forms link back to the physical privacy concerns of having one's own space free from intrusion (Kaminski, 2015; Kaminski et al., 2017). However, the notion of boundary management is broader than "freedom from" surveillance, as it understands privacy as a protection of individuals' agency to make their own life choices and thus ultimately as "freedom from unreasonable constraints on the construction of identity" (Carnevale, 2016, p. 147). Social informational privacy concerns rest on the ability of a user to understand how information shared with the social robot is processed, especially considering the anthropomorphic effect of social robots. This effect has been widely recognized in the literature on social robots (Darling, 2016; Weiss et al., 2009), also as an important aspect regarding privacy (Calo, 2012; Syrdal et al., 2007). Studies in the field of human-robot-interaction have shown that humans tend to anthropomorphize or zoomorphize social robots (Fong et al., 2003). This increases their pervasiveness compared with other connected technology (Turkle, 2011), in the sense that humans might feel more inclined to see the robots as companions or friends. In turn, they will be more likely to entrust the robots with personal, potentially sensitive information. Social informational privacy thus includes aspects of boundary management between a social robot and a user (e.g., can secret information such as passwords be revealed to the robot), touching on aspects of interdependency and bonding (Calo, 2012). However, social informational privacy concerns not only relate to the interaction between the user and the robot itself but also to the interaction between individuals through a robot, for example when a robot is hacked or surveillance takes place through a telepresence robot.

Empirical Research on Privacy Concerns in the Context of Social Robots

As mentioned above, empirical research on privacy concerns in the context of social robots exists but remains in its infancy. For example, an exploratory study based on qualitative interviews investigated privacy perceptions of the social workplace robot Snackbot (Lee et al., 2011). The interviews revealed that most participants were not able to grasp the types of data Snackbot collects and failed to differentiate between sensed data ("what the robot sees/hears") and inferred information ("what the robot knows," p. 182). The authors also found that Snackbot's anthropomorphic form could mislead participants in their understanding of the capabilities to record information. Specifically, participants did not expect that Snackbot could sense objects or people behind it.

Other surveys explored the issue of information disclosure in human-robot interactions. In a study conducted by Syrdal et al. (2007), participants' fear of robots storing and accessing sensitive information about individuals' behavior was considered a "necessary evil" that had to be tolerated so long as the social robots brought them benefits. Similarly, Butler et al. (2015) analyzed the privacy-utility tradeoff for teleoperated robots, with the aim of reducing "the quantity or fidelity of visual information received by the teleoperator to preserve the end-user's privacy" (p. 27), while still providing sufficient information for the robot to be able to fulfill its tasks. The authors provide a framework for understanding what visual filters may be applied to balance the privacy needs of the participants with the information needed to perform actions by the teleoperator. Studies on telepresence robotics have also been conducted with non-academic participants. Krupp et al. (2017) carried out in-depth focus groups (13 participants, 3 sessions, 2 hours long) discussing privacy in telepresence robotics. Privacy concerns expressed by the participants ranged between fear of hackers infiltrating the systems, fear of constant monitoring and recording of embarrassing moments, and fears of becoming prey to even more personalized marketing practices.

In addition, some general population surveys have assessed citizens' attitudes toward robots, including potential concerns (e.g., Eurobarometer, 2012, 2015; Madden & Rainie, 2015). While a majority of respondents in the European Union had a positive opinion of robots (64%), a common fear was robots stealing people's jobs (70%). Moreover, a majority of respondents felt uncomfortable with the thought of robots providing companionship to older people and with robots being used for surgeries on them personally (Eurobarometer, 2015). The latter findings might be connected to perceptions of privacy, although the survey did not explicitly ask for privacy concerns. In the US, Liang and Lee (2017) investigated individuals' fear of robots and artificial intelligence based on national representative data. They found that about one fourth of the population had heightened fear levels and that fear of robots and artificial intelligence was positively correlated with other types of fear, including fear of government drone use.

Across this literature, few quantitative and survey-based studies have assessed privacy concerns, their antecedents and outcomes (Lutz et al., 2019). Thus, we lack knowledge about whether established privacy theories could prove useful for social robots. We also know little about non-expert opinions and concerns about social robots. This lack of knowledge could be problematic, as social robots are sometimes introduced without a thorough assessment of potential user concerns. Our contribution tries to overcome some of these gaps.

Trust and Its Link to Privacy

Trust is a complex phenomenon and so is its link to privacy (Waldman, 2018). On the one hand, and from a social informational privacy perspective, privacy allows psychological release functions and enables interpersonal relationships that are built upon trust and trusting beliefs (Tamò-Larrieux, 2018; Westin, 1967). On the other hand, and from an institutional privacy perspective, privacy features of services and products enhance consumer trust in the provider, which in turn is a key element for economic success (Hartzog, 2018; Tamò-Larrieux, 2018). In our survey, we rely on the conventional definition of trust as "a psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or behaviors of another" (Rousseau et al., 1998, p. 395). This definition aligns with Möllering's (2001) conceptualization of trust as a three-step mental process of expectation, interpretation, and suspension. The release of private information to a social robot requires trust, as such a release requires a favorable expectation of an outcome that is uncertain. The interpretation of whether the outcome will be favorable or not can rely on various rational and emotional elements (or a mix thereof); in the end, a user must have enough "good reasons" to trust to interact with a social robot. Möllering (2001) calls the moment in which the interpretation becomes accepted and the unknowable momentarily certain "suspension." Suspension represents an element of faith toward the outcome and "enables the leap of trust" (Möllering, 2001, p. 414).

Given its importance, policymakers are trying to establish trust in new technologies by enacting ethical guidelines. In particular, the rise of artificial intelligence (AI) has pushed the European Commission and the global community to establish ethical guidelines that promote trustworthy AI (Delcker, 2019; European Commission, 2018). Social robots are and will be equipped with AI-systems, which is why the adherence to ethical principles (e.g., respect for human autonomy, fairness, explicability, prevention of harm) of AI likewise affects the development of social robots. We consider privacy and data protection important aspects of ethical and trustworthy technology but cannot fully do justice to this emerging literature in ethics and AI here.

Model and Theoretical Development

Based on our discussion of the privacy literature above, as well as adjacent work on technology acceptance and trust in the context of robots (Alaiad & Zhou, 2014), we propose the following research model (Figure 1).

In our model, behavioral intention is the key dependent construct. We did not include actual behavior because few of our respondents could be expected to have experience in interacting with robots, thus making behavioral assessments unreliable and speculative. Attitudes are represented by privacy concerns, trusting beliefs, perceived benefits of robots, and scientific interest. Social psychological theories, such as the theory of planned behavior (TPB), have stressed the importance of social influence in explaining behavioral intentions. Social influence, or subjective norm (the two are often used synonymously), describes the "perceived social pressure to perform or not to perform the behavior" (Ajzen, 1991, p. 188). As social robots are employed in social settings, we included social influence as an

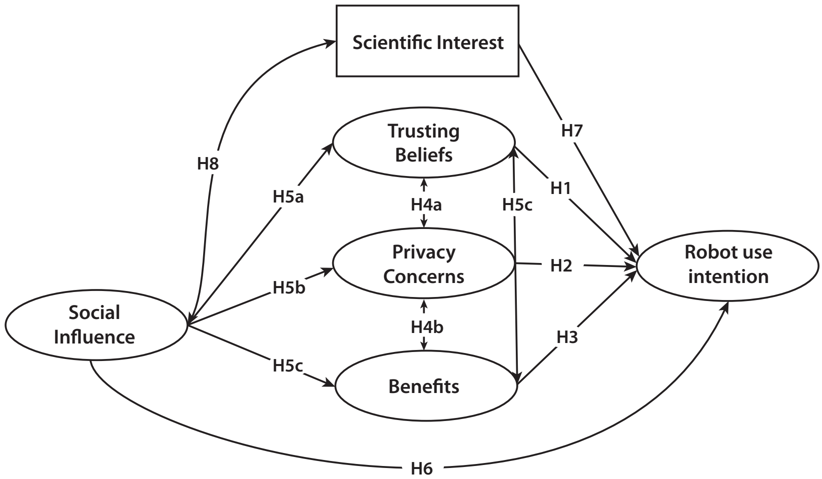


FIGURE 1 Research Model

independent construct in the model affecting use intention, trusting beliefs, privacy concerns, and perceived benefits. Finally, we included scientific interest as an attitudinal control variable affecting the intention to use social robots and being itself influenced by social influence. In the following paragraphs, we explain the model and present the hypotheses.

Among the attitudinal constructs, trust, and more specifically trusting beliefs, should be associated with robot use intentions. Trusting beliefs contain several subdimensions, the most important of which are a trustor's competence, benevolence, and integrity beliefs (McKnight et al., 2002). In other words, to form trust, the trustor must think that the trustee is competent, benevolent, and honest. If this is the case, individuals are more likely to develop trusting intentions, which will eventually result in a certain trusting behavior such as technology adoption. Based on this mechanism established in the trust literature, we propose the following hypothesis. We focus on trusting beliefs and the distinction of competency, benevolence, and integrity because such a conceptualization is easier to operationalize than, for example, Möllering's (2001) leap of trust approach.

H1: *Trusting beliefs in robot manufacturers have a positive effect on robot use intentions.*

Overcoming privacy concerns seems to be an initial requirement for the intention to start using social robots. When citizens perceive the privacy risks of social robots to be high or when they have had adverse experiences with social robots, we expect their intention to adopt them to decrease. At the same time, extensive research on online privacy and self-disclosure has found that individuals' privacy attitudes—including concerns—are often not in line with their behavior (Kokolakis, 2017). Despite substantial privacy concerns, many users of social media and other online services disclose a lot of sensitive information and engage superficially in privacy protection behavior. This misalignment between online privacy attitudes and behavior has been termed the "privacy paradox" (Barnes, 2006). A range of empirical studies has found a privacy paradox but some studies, especially newer ones, reported significant effects between privacy attitudes and behavior, thus rejecting the notion of the privacy paradox. Kokolakis offers a systematic assessment of this literature and shows the inconclusive empirical evidence. Given the novelty of the topic (social robots), current low adoption rates, and our focus on intention rather than behavior as the dependent variable, we would expect privacy considerations to have a significant effect on use intention.

H2: *Privacy concerns about robots have a negative effect on robot use intentions.*

As shown in the literature review, we identify three key aspects of privacy—physical privacy, institutional informational privacy, and social informational privacy—that we apply to the context of social robots (summarized in Table 1). We check how each type affects use intention differently. We would expect that more familiar concerns might deter people more from robots than unfamiliar or even intangible concerns. In that sense, many citizens will be familiar with informational privacy concerns (both institutional and social) through their Internet and social media use, but will have limited familiarity with robots' physical risk potential. Thus, we expect differentiated roles of the three privacy concerns considered.

TABLE 1 Overview of Privacy Aspects Considered

Privacy aspects	Description
Physical privacy	Privacy considerations that revolve around non-interference by social robots themselves and their interaction with physical objects and spaces (e.g., by entering private rooms and using personal objects).
Institutional informational privacy	Privacy considerations that revolve around information control and data protection from and data collection practices by institutions, in particular social robot manufacturers, government agencies, and third parties (data brokers, cloud storage providers).
Social informational privacy	Privacy considerations that revolve around access to a person and data protection from and data collection practices by individuals (e.g., other users, hackers).

Within the privacy paradox literature, the privacy calculus is the dominant theoretical explanation (Dinev & Hart, 2006). According to this approach, users perform a mental calculus weighing the risks and benefits of an online technology against each other. If the benefits outweigh the risks, they will start or keep using the technology. In a similar vein, if social robots are perceived as extremely useful for someone's personal life, individuals will be more likely to develop use intentions, despite potential privacy concerns. In our case, we considered two key benefits of social robots: functional benefits (Lin, 2012) and emotional benefits (Yu et al., 2015). For the analysis we consider them in conjunction. Based on previous research (Alaiad & Zhou, 2014) and one of the core premises of TPB, we expect perceived benefits to have pronounced and comparatively strong effects on use intentions.

H3: *Perceived benefits of social robots have a positive effect on social robot use intentions.*

In our model, privacy concerns, trusting beliefs, and perceived benefits are all conceptualized as attitudes. We understand the relationship between these constructs as correlational associations rather than causal effects. Previous research in different online contexts has specified the relationship between privacy and trust in both directions. Privacy concerns can lower trust in a service provider (Bart et al., 2005; Beldad et al., 2010; Hoffmann et al., 2014) but lowered levels of trust might also result in heightened privacy concerns (Krasnova et al., 2010). In our case, we would argue that privacy concerns can lower trust in a social robot manufacturer. If users or potential users worry that a social robot manufacturer cannot protect their data and privacy to a sufficient extent, they will potentially challenge the manufacturer's competence, benevolence, and integrity, thus having lowered trust. At the same time, trusting beliefs might decrease privacy concerns or the two might mutually enforce each other. Similarly, an increase in privacy concerns might result in the perceived benefits being less salient and a mental reconfiguration of the perceived benefits might affect someone's privacy concerns. Finally, we think that the perceived benefits and trust go hand in hand as well, leading us to hypotheses H4a–H4c.

H4a: *Privacy concerns about social robots correlate negatively with trusting beliefs in social robot manufacturers.*

H4b: *Privacy concerns about social robots correlate negatively with perceived benefits of social robots.*

H4c: *Trusting beliefs about social robots correlate positively with perceived benefits of social robots.*

As mentioned, social influence, or subjective norm, describes the "perceived social pressure to perform or not to perform the behavior" (Ajzen, 1991, p. 188). Thus, it refers to the social environment and its expectations toward an individual. Since trusting beliefs, privacy concerns, and perceived benefits of social robots do not form in a social vacuum, we hypothesize that they are all affected by social influence. More specifically, we expect an encouraging and technology-affine social environment to enhance trust, reduce privacy concerns, and make the perceived benefits of social robots more salient.

H5a: *Social influence has a positive effect on trusting beliefs about social robots.*

H5b: *Social influence has a negative effect on privacy concerns about social robots.*

H5c: *Social influence has a positive effect on perceived benefits of social robots.*

Social influence or subjective norm has proven to be an important predictor of behavioral intention in TPB (McEachan et al., 2011). Similarly, theories of technology adoption have stressed the importance of social factors, for example within the technology acceptance model framework (Venkatesh & Morris, 2000; Venkatesh, Morris, Davis, et al., 2003). In this understanding, social influence should enhance individuals' behavioral intention to adopt a new technology. Social robots, as a technology that is not yet widely adopted, will likely be adopted earlier when someone's network expects and encourages their use. Thus, citizens who are part of more social robot-friendly communities will have higher intentions to use them.

H6: *Social influence has a positive effect on social robot use intentions.*

We included scientific interest as a control variable. More scientifically interested citizens will be more up-to-date with current technological developments, also regarding social robots. As social robots are still not a mainstream technology, we assessed scientific interest as a proxy for awareness and knowledge of social robot technology. The rationale for a positive effect is that scientifically interested citizens will be better able to assess the benefits and risks of the technology, including the privacy risks. They might also be more technologically curious and open-minded, having higher willingness to try out social robots despite a lack of widespread adoption. In that sense and based on diffusion of innovation theory (Rogers, 2003), scientifically interested citizens should be more likely to be early adopters of social robots. By including scientific interest, we follow other survey-based studies (Eurobarometer, 2012).

H7: *Scientific interest has a positive effect on social robot use intentions.*

Finally, as scientific interest depends on the social milieu and environment, we hypothesize that these variables will positively correlate with each other. Again, we specify this relationship as a correlational association rather than a causal one. A social environment that is positive toward social robots and encourages their use might stimulate someone's general scientific interest. Similarly, according to homophily theory (McPherson et al., 2001), scientifically interested individuals might prefer a social environment that is positive toward social robots.

H8: *Scientific interest and social influence correlate positively.*

Methods

To test the model in Figure 1, we used a survey-based approach. We think that surveys are a useful tool to assess individuals' perceptions and beliefs, allowing for descriptive and correlational analyses. Moreover, in a systematic literature on privacy in the context of robots, the authors found that very few studies rely on surveys, so that limited evidence about privacy attitudes and concerns is available from a quantitative perspective (Lutz, Schöttler, et al., 2019).

Sample

We rely on data collected through a survey on Amazon Mechanical Turk (MTurk) in June 2016. Participants were all residents in the United States (US).[1] They were offered a monetary incentive of 2 US Dollars and survey completion took 15 minutes on average. Thus,

1 We are conscious that our position as European researchers might result in cultural bias in the interpretation of data collected in the US. However, such bias is probably mitigated by our extensive collaboration with US-based researchers, by our experience with analyzing US-based data across several projects, and by having spent considerable time at US institutions through research stays, conferences, and workshops. Despite not removing cultural bias entirely, we hope that our familiarity with the US context and culture has reduced inherent bias.

the average hourly wage for filling out the survey was approximately 8 US Dollars. 501 respondents started the survey, 480 of whom are included in the structural equation model and had no or very few missing values. 54.5% of the respondents were male, 45% female, and 0.5% (two respondents) identified as other. The respondents were relatively highly educated, with 35% having some college education, 38% a bachelor's degree, 8% a master's degree, and 2% a doctorate. Only 16% had a high school degree as their highest degree and 1% responded with "other." The average age was 34 (median 32), with a range between 18 and 74 years and a standard deviation of 10.5 years. Thus, the sample is not representative of the US general population or US adult population. The questionnaire was aimed at non-experts to capture the general privacy concerns associated with social robots.

Measurement

To make the questions relatable and prime the respondents to answer the questions for social robots (rather than industrial and other non-social robots), we showed pictures of different social robots interacting with people at the very beginning of the survey (Appendix A). The wording of all items is shown in Appendix B.

We used four items to measure respondents' intention to use social robots. A sample item was: *"I would very much like to have such a robot at home."* The scales used to measure trusting beliefs (McKnight et al., 2002) and social influence (Venkatesh, Morris, Davis, et al., 2003) were derived from well-established models. They were adapted to the context of social robots. The measures for perceived benefits were taken from the Special Eurobarometer 382 and 427 surveys on public attitudes toward robots (Eurobarometer, 2012, 2015). Scientific interest was measured with one item from Eurobarometer (2012). The measures for informational and global privacy concerns were based on previous studies (Malhotra et al., 2004; Stutzman et al., 2011), but adapted to the context of social robots. Within the global privacy concerns scale, the first item (*"Overall, I see a real threat to my privacy due to the robot."*) assesses privacy in particular, while the remaining three items capture concerns more broadly. Nevertheless, all items load neatly on one factor, with good reliability and convergent validity values (Cronbach's α of 0.90, and average variance extracted—AVE—of 0.71). Thus, we opted to retain this scale instead of relying on a less robust single-item measurement. Within the seven informational privacy concerns items, three items refer to social informational privacy concerns and four items to institutional informational privacy concerns. The scale for physical concerns was self-developed because we did not encounter studies which contained such a scale. However, the question prompt was adapted from Stutzman et al. (2011). Physical privacy concerns were measured with five items.

We relied on 5-point Likert scales ranging from "strongly disagree" to "strongly agree" for all items, except for privacy concerns. Here, respondents could assess their concern on a 5-point scale ranging from "no concern at all" (1) to "very high concern" (5). All scales reveal good measurement properties in terms of internal consistency, reliability, and validity. The measurement model (Appendix C, Table B) thus satisfies the necessary conditions to report the structural model, displaying both convergent and discriminant validity (Bollen, 1989; Fornell & Larcker, 1981; Netemeyer et al., 2003). As the only exception for discriminant validity, the squared correlation between perceived benefits and intentions exceeds the AVE of benefits by 0.01 (Appendix C, Table C). We decided to keep these two constructs as separate because of their theoretical importance, distinct conceptualization, and because of the correlation being only very little above the threshold.

Methodological Approach

We relied on structural equation modeling (SEM) to test the research model, combining advantages of confirmatory factor analysis and regression analysis. We used robust maximum-likelihood estimation (MLR) in MPlus (Version 7) to account for non-normality, heteroscedasticity, and other possible sources of distortion (Byrne, 2012). All models reported had sufficient goodness of fit indices, with the overall privacy concerns model being the least good: Chi-Square = 508.4; degrees of freedom = 213; Root mean square error of approximation (RMSEA) = 0.054; Comparative fit index (CFI) = 0.95; Tucker-Lewis index (TLI) = 0.94; Standardized root mean square residuals (SRMR) = 0.045.

Results

Before turning to the structural model, we present descriptive results. As outlined before, we distinguish three types of privacy concerns: *physical privacy concerns*, *institutional informational privacy concerns*, and *social informational privacy concerns*. In addition, we included a measure for overall privacy concerns. In the following comparisons, it is important to remember that the wording of the different privacy constructs varies from more moderate to more extreme, depending on the risk described. Thus, the comparisons have to be interpreted carefully. The arithmetic means for *physical privacy concerns* range from 1.90 (the social robot asking personal questions), to 2.56 (the social robot damaging or dirtying personal belongings), with a global average of 2.22. This indicates low concern. The arithmetic means for *institutional informational privacy concerns* range from 3.70 (insufficient data protection), to 3.74 (selling data), with a global average of 3.72. This indicates high concern. The arithmetic means for *social informational privacy concerns* range from 2.93 (stalking), to 3.53 (hacking), with a global average of 3.17. This indicates moderate concern. Finally, the global privacy concerns measure lies in between informational privacy concerns on the one hand and physical privacy concerns on the other hand, with a global average of 2.61.

The findings indicate that the respondents are most concerned about institutional aspects of privacy (i.e., data protection on the side of the manufacturers). They seem to be unconcerned about physical privacy. However, they are somewhat concerned about other users using the social robot for malicious purposes such as stalking or hacking. Overall, respondents have moderate privacy concerns about social robots.

The model for physical privacy concerns is shown in Figure 2. Trusting beliefs and privacy concerns have no significant effect on robot use intention, rejecting H1 and H2. However, perceived benefits have a positive influence on social robot use intention, supporting H3. Physical privacy concerns, trusting beliefs and perceived benefits correlate significantly and in the expected direction with each other, thus supporting H4. Physical privacy concerns are not affected by social influence, but social influence has the predicted effect on trusting beliefs and perceived benefits, partially supporting H5. Social influence has a positive and significant effect on robot use intention, supporting H6. Finally, scientific interest has no significant effect on robot use intention, rejecting H7, but is itself positively affected by social influence, supporting H8.

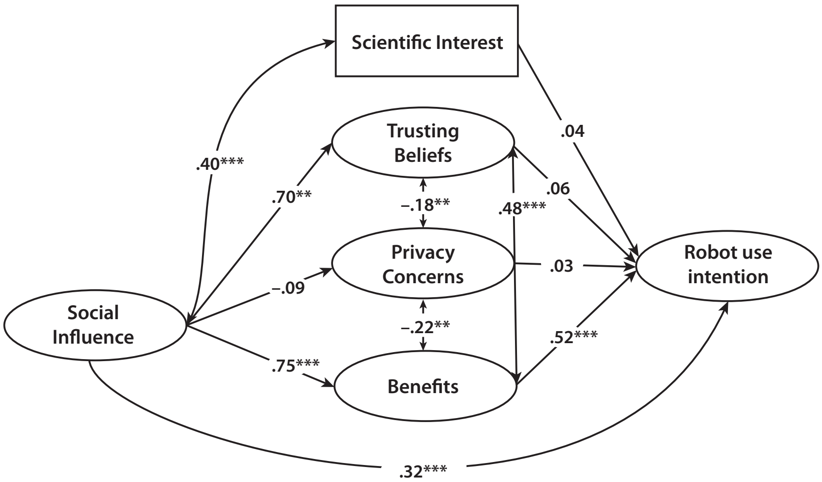


FIGURE 2 Physical Privacy Concerns Model

The models for institutional and social informational privacy concerns (Figures 3 and 4) are vastly similar to the physical privacy concerns model. However, both forms of informational privacy concerns do not correlate significantly with trusting beliefs and benefits, so that H4 is partly supported.

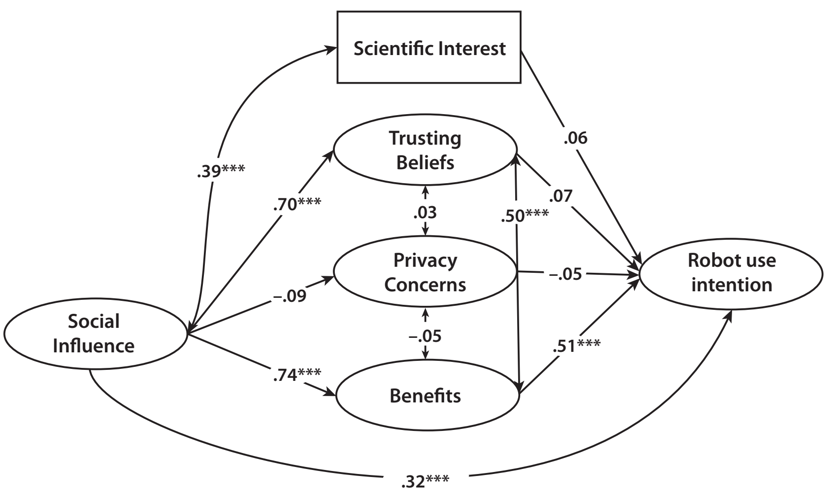


FIGURE 3 Institutional Informational Privacy Concerns Model

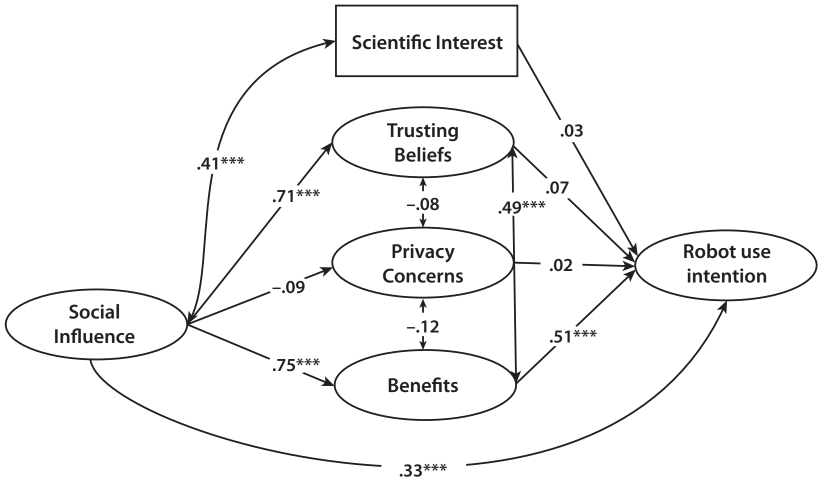


FIGURE 4 Social Informational Privacy Concerns Model

Turning to the model with global privacy concerns, we find vastly similar effects again (Figure 5), except for the role of social influence, which has the expected effect on all constructs. Therefore, H5, H6, and H8 are supported.

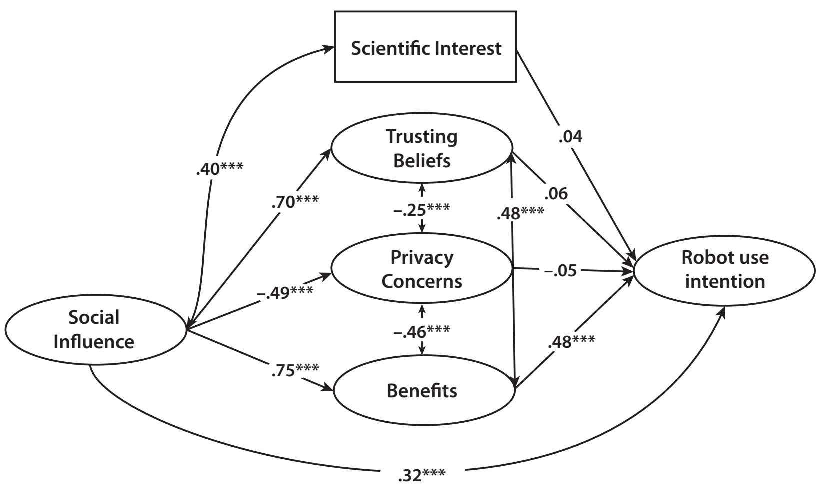


FIGURE 5 Overall Privacy Concerns Model

Across the models, we were able to explain between 72% (overall privacy concerns) and 73% (other privacy concern types) of the variance in social robot use intention. Thus, the small number of constructs has high predictive power. Particularly, the combination of perceived benefits and social influence seems to be able to predict social robot use intention strongly.

Discussion and Conclusion

Early studies on online privacy have focused on institutional privacy threats such as how service providers handle user data, especially in the domain of e-commerce (Jarvenpaa et al., 1999; McKnight et al., 2002). The emergence of social media has further intensified the debate on online privacy (Ellison et al., 2007; Krasnova et al., 2010). Accordingly, with social media, institutional privacy concerns are compounded by social privacy concerns: concerns about privacy threats that are caused by other users rather than service providers or third-party institutions (Raynes-Goldie, 2010; Young & Quan-Haase, 2013). We argue that social robots add yet another layer, with entirely new challenges to privacy, as they have mobility and thus access to private rooms (Calo, 2012).

Previous research has found a paradoxical disparity between users' privacy concerns and their online behaviors, such as a lack of privacy protection and a willingness to engage in extensive data sharing (Chen & Rea, 2004; Milne et al., 2009). Based on these findings, we developed a nomological model that considers distinct explanations for users' social robot use intention despite privacy concerns. We distinguished three types of privacy concerns and found that they were unequally pronounced. Respondents worried most about their informational privacy, especially with regard to institutions such as the social robot manufacturer. Social privacy risks, such as hacking and stalking, also evoked considerable concerns. Physical privacy concerns were less prevalent.

Perceived benefits and social influence had a significant and positive effect on social robot use intention. Some forms of privacy concerns were themselves significantly affected by social influence. This points to the explanatory value of including the social environment when looking at social robot adoption.

Our study provides a number of theoretical and practical *implications*. First, we established the existence of compounded privacy concerns in the social robot context, as we found evidence of both informational and physical privacy concerns. Second, we found a privacy paradox in line with previous studies (Kokolakis, 2017), as we detected that neither informational nor physical privacy concerns significantly affected social robot use intention. Third, we found that social dynamics are especially important in the analysis of social robot use intention. In fact, social influence drove intentions in three ways. First, it directly increased social robot use intentions; second, it reduced respondents' concerns, at least in the overall model; third and finally, it strongly increased the perceived benefits of social robots. Together, these findings demonstrate that social norms are of crucial importance in the context of social robots. As such, robot manufacturers would do well to invest in community management and they should rely heavily on word-of-mouth promotion.

As our study does not illuminate the concerns of experienced users (we sampled individuals not familiar with social robots for the most part), the implications of this research for social robot manufactures are not entirely clear-cut. Despite the apparent privacy paradox, recent media coverage of privacy issues with Internet-of-things devices, such as toys, indicates increasing public attention to these matters (Mathews, 2017). The effects of such public debate could affect the adoption of social robots as privacy concerns may influence the trust of users in the social robot manufactures and thus have an effect on the use intentions.

Thus, social robot manufacturers should be aware of the fact that consumers might value privacy and consider it in their purchasing decisions when faced with tangible risks. In that sense, manufacturers might want to increase investments into privacy-sensitive robotics (Rueben, Aroyo, et al., 2018). Not only should manufacturers develop social robots that are privacy-friendly but they should also communicate their privacy-protection efforts to potential customers in concise and transparent ways (Felzmann et al., 2019). Overall, this study highlights the compounded privacy challenges that are associated with social robots and points to its differentiated nature in affecting social robot use intention. Even if survey results on social robots and privacy concerns are bound to be abstract due to a still limited daily interaction with social robots in households, schools, or at work, empirical findings about privacy can be helpful for different stakeholders, from the academic community to practitioners and regulatory bodies.

Limitations

Our research comes with several limitations which may inspire future research on the topic. First, we conducted a cross-sectional study with a relatively low number of participants. Thus, future research should use larger and longitudinal samples, if possible representative of the whole population. Moreover, it should compare owners and users of social robots with those who are unfamiliar with them in terms of privacy concerns to investigate experience effects. Second, for the sake of brevity, our questionnaires did not assess social robots' characteristics and their perception. Future research might work with field and lab experiments and use a systematic variation of social robot characteristics to assess privacy concerns with social robots more broadly. In that regard, surveys on social robots with non-users are bound to stay relatively abstract. Thus, the results might differ from a controlled lab setting where users get to experience social robots firsthand. However, previous research has indicated that research on non-experts, such as ours, can be helpful to assess individuals' attitudes and fears of social robots, even if they have not used such technology themselves (Liang & Lee, 2017). Third, we could not assess contextual characteristics, such as users' cultural backgrounds or their social milieus. Future research could delve deeper into user characteristics and users' composition of social networks to achieve a more holistic understanding of privacy.

Acknowledgments

We would like to thank three anonymous peer reviewers of this article and the volume editor, Prof. Leopoldina Fortunati, for very helpful feedback and guidance during the peer review process.

The Research Council of Norway funded the data collection for this article as part of the first author's research funding under the Grant Agreements "247725 Fair Labor in the Digitized Economy" and "275347 Future Ways of Working in the Digital Economy."

Author Biographies

Christoph Lutz (Dr./PhD, University of St. Gallen) is an Associate Professor at the Nordic Centre for Internet and Society, BI Norwegian Business School (Oslo, Norway). His research interests include online participation, privacy, the sharing economy and social robots. Christoph has published widely in top-tier journals in this area.

https://orcid.org/0000-0003-4389-6006

Aurelia Tamò-Larrieux (Dr./PhD, University of Zurich) is a Postdoctoral Researcher at the Center for Information Technology, Society, and Law (ITSL) and Digital Society Initiative at the University of Zurich (Switzerland). Her research interests include privacy, especially privacy-by-design, data protection, social robots, and artificial intelligence. Aurelia has published widely in this area, including the book *Designing for Privacy and its Legal Framework* (Springer, 2018).

https://orcid.org/0000-0003-3404-7643

References

Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes, 50*(2), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

Alaiad, A., & Zhou, L. (2014). The determinants of home healthcare robots adoption: An empirical investigation. *International Journal of Medical Informatics, 83*(11), 825–840. https://doi.org/10.1016/j.ijmedinf.2014.07.003

Altman, I. (1975). *The Environment and Social Behavior—Privacy, Personal Space, Territory, Crowding*. Monterey: Wadsworth Publishing Company.

Barnes, S. B. (2006). A privacy paradox: Social networking in the United States. *First Monday, 11*(9). https://doi.org/10.5210/fm.v11i9.1394

Bart, Y., Shankar, V., Sultan, F., & Urban, G. L. (2005). Are the drivers and role of online trust the same for all web sites and consumers? A large-scale exploratory empirical study. *Journal of Marketing, 69*(4), 133–152. https://doi.org/10.1509/jmkg.2005.69.4.133

Beldad, A., De Jong, M., & Steehouder, M. (2010). How shall I trust the faceless and the intangible? A literature review on the antecedents of online trust. *Computers in Human Behavior, 26*(5), 857–869. https://doi.org/10.1016/j.chb.2010.03.013

Bollen, K. A. (1989). *Structural equations with latent variables*. Wiley Interscience.

Butler, D. J., Huang, J., Roesner, F., & Cakmak, M. (2015). The privacy-utility tradeoff for remotely teleoperated robots. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 27–34). ACM. https://doi.org/10.1145/2696454.2696484

Bygrave, L. A. (2002). *Data protection law: Approaching its rationale, logic and limits*. Wolters Kluwer.

Byrne, B. (2012). *Structural equation modeling with Mplus*. Routledge.

Calo, R. (2012). Robots and privacy. In P. Lin, G. Bekey, & K. Abney (Eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (pp. 187–202). MIT Press.

Calo, R. (2016). Robots in American law. *University of Washington School of Law Research Paper*, No. 2016-04.

Carnevale, A. (2016). Will robots know us better than we know ourselves? *Robotics and Autonomous Systems, 86*, 144–151. https://doi.org/10.1016/j.robot.2016.08.027

Chen, K., & Rea, A. I. (2004). Protecting personal information online: A survey of user privacy concerns and control techniques. *Journal of Computer Information Systems, 44*(4), 85–92. https://doi.org/10.1080/08874417.2004.11647599

Darling, K. (2016), Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior toward robotic objects. In R. Calo, M. Froomkin & I. Kerr. (Eds.), *Robot Law* (pp. 213–234). Edward Elgar.

Delcker, J. (2019). *US to endorse new OECD principles on artificial intelligence*. *Politico*, May 19, 2019. Archived at https://web.archive.org/web/20190520045733/https://www.politico.eu/article/u-s-to-endorse-new-oecd-principles-on-artificial-intelligence/

Dinev, T., & Hart, P. (2006). An extended privacy calculus model for e-commerce transactions. *Information Systems Research, 17*(1), 61–80. https://doi.org/10.1287/isre.1060.0080

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook 'friends': Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication, 12*(4), 1143–1168. https://doi.org/10.1111/j.1083-6101.2007.00367.x

Eurobarometer. (2012). *Special Eurobarometer 382: Public attitudes toward robots*. http://ec.europa.eu/public_opinion/archives/ebs/ebs_382_en.pdf

Eurobarometer. (2015). *Special Eurobarometer 427: Autonomous systems*. http://ec.europa.eu/public_opinion/archives/ebs/ebs_427_en.pdf

European Commission. (2018). *Ethics guidelines for trustworthy AI: High-Level Expert Group on Artificial Intelligence*. Archived at https://web.archive.org/web/20190423110004/https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai

Felzmann, H., Fosch-Villaronga, E., Lutz, C., & Tamò-Larrieux, A. (2019). Robots and Transparency: The Multiple Dimensions of Transparency in the Context of Robot Technologies. *IEEE Robotics & Automation Magazine, 26*(2), 71–78. https://doi.org/10.1109/MRA.2019.2904644

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and autonomous systems, 42*(3), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Gupta, S. K. (2015, 8 September). Six recent trends in robotics and their implications. *IEEE Spectrum*. Archived at https://web.archive.org/web/20190801112252/https://spectrum.ieee.org/automaton/robotics/home-robots/six-recent-trends-in-robotics-and-their-implications

Hartzog, W. (2018). *Privacy's blueprint: The battle to control the design of new technologies*. Harvard University Press.

Hoffmann, C. P., Lutz, C., & Meckel, M. (2014). Digital natives or digital immigrants? The impact of user characteristics on online trust. *Journal of Management Information Systems, 31*(3), 138–171. https://doi.org/10.1080/07421222.2014.995538

Jarvenpaa, S. L., Tractinsky, N., & Saarinen, L. (1999). Consumer trust in an Internet store. *Journal of Computer-Mediated Communication, 5*(2), 34–67. https://doi.org/10.1111/j.1083-6101.1999.tb00337.x

Kaminski, M. E. (2015). Robots in the home: What will we have agreed to? *Idaho Law Review, 51*, 661–677. Archived at https://www.uidaho.edu/-/media/UIdaho-Responsive/Files/law/law-review/articles/volume-51/51-3-kaminski-margot-e.pdf

Kaminski, M. E., Rueben, M., Grimm, C., & Smart, W. D. (2017). Averting robot eyes. *Maryland Law Review, 76*, 983–1023. https://ssrn.com/abstract=3002576

Kokolakis, S. (2017). Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & Security, 64*, 122–134. https://doi.org/10.1016/j.cose.2015.07.002

Koops, B. J., Newell, B. C., Timan, T., Škorvánek, I., Chokrevski, T., & Galič, M. (2016). A typology of privacy. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2754043

Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: Why we disclose. *Journal of Information Technology, 25*(2), 109–125. https://doi.org/10.1057/jit.2010.6

Krupp, M. M., Rueben, M., Grimm, C. M., & Smart, W. D. (2017, March). Privacy and telepresence robotics: What do non-scientists think? In *Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 175–176). ACM. https://doi.org/10.1145/3029798.3038384

Lee, M. K., Tang, K. P., Forlizzi, J., & Kiesler, S. (2011, March). Understanding users' perception of privacy in human-robot interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 181–182). ACM.

Liang, Y., & Lee, S. A. (2017). Fear of autonomous robots and artificial intelligence: Evidence from national representative data with probability sampling. *International Journal of Social Robotics, 9*(3), 379–384. https://doi.org/10.1007/s12369-017-0401-3

Lin, P. (2012). Introduction to Robot Ethics. In P. Lin, G. Bekey, & K. Abney (Eds.), *Robot Ethics: The Ethical and Social Implications of Robotics* (1st ed., pp. 3–16). MIT Press.

Lutz, C., Schöttler, M., & Hoffmann, C. P. (2019). The privacy implications of social robots: Scoping review and expert interviews. *Mobile Media & Communication*, *7*(3), 412–434.

Lutz, C., & Tamò, A. (2015). RoboCode-Ethicists: Privacy-friendly robots, an ethical responsibility of engineers? In *Proceedings of the 2015 ACM Web Science Conference*, Oxford UK, 28 June–1 July 2015. ACM. https://doi.org/10.1145/2786451.2786465

Lutz, C., & Tamò, A. (2018). Communicating with robots: ANTalyzing the interaction between healthcare robots and humans with regards to privacy. In A. Guzman (Ed.), *Human-Machine Communication: Rethinking Communication, Technology, and Ourselves* (pp. 145–165). Peter Lang.

Madden, M., & Rainie, L. (2015). Americans' attitudes about privacy, security and surveillance. Pew Internet, Science & Tech Report. Archived at https://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/

Malhotra, N. K., Kim, S. S., & Agarwal, J. (2004). Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information Systems Research, 15*(4), 336–355. https://doi.org/10.1287/isre.1040.0032

Mathews, L. (2017, 28 February). The latest privacy nightmare for parents: Data leaks from smart toys. *Forbes*. Archived at https://www.forbes.com/sites/leemathews/2017/02/28/cloudpets-data-leak-is-a-privacy-nightmare-for-parents-and-kids

McEachan, R. R. C., Conner, M., Taylor, N. J., & Lawton, R. J. (2011). Prospective prediction of health-related behaviours with the theory of planned behaviour: A meta-analysis. *Health Psychology Review, 5*(2), 97–144. https://doi.org/10.1080/17437199.2010.521684

McKnight, D. H., Choudhury, V., & Kacmar, C. (2002). Developing and validating trust measures for e-commerce: An integrative typology. *Information Systems Research, 13*(3), 334–359. https://doi.org/10.1287/isre.13.3.334.81

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology, 27*(1), 415–444. https://doi.org/10.1146/annurev.soc.27.1.415

Milne, G. R., Labrecque, L. I., & Cromer, C. (2009). Toward an understanding of the online consumer's risky behavior and protection practices. *Journal of Consumer Affairs*, *43*(3), 449–473. https://doi.org/10.1111/j.1745-6606.2009.01148.x

Möllering, G. (2001). The nature of trust: From Georg Simmel to a theory of expectation, interpretation and suspension. *Sociology, 35*(2), 403–420. https://doi.org/10.1177/S0038038501000190

Netemeyer, R. G., Bearden, W. O., & Sharma, S. (2003). *Scaling procedures. Issues and applications*. Sage.

Pavlou, P. A. (2011). State of the information privacy literature: where are we now and where should we go? *MIS Quarterly, 35*(4), 977–988. https://doi.org/10.2307/41409969

Raynes-Goldie, K. (2010). Aliases, creeping, and wall cleaning: Understanding privacy in the age of Facebook. *First Monday, 15*(1). https://doi.org/10.5210/fm.v15i1.2775

Rogers, E. (2003). *Diffusion of innovations* (4th ed.). Free Press.

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Not so different after all: A cross-discipline view of trust. *Academy of Management Review, 23*(3), 393–404. https://doi.org/10.5465/amr.1998.926617

Rueben, M., Aroyo, A. M., Lutz, C., Schmölz, J., Van Cleynenbreugel, P., Corti, A., Agrawall, S., & Smart, W. D. (2018). Themes and research directions in privacy sensitive robotics. In *2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)* (pp. 1–8). IEEE. https://doi.org/10.1109/arso.2018.8625758

Rueben, M., Grimm, C. M., Bernieri, F. J., & Smart, W. D. (2017). A taxonomy of privacy constructs for privacy-sensitive robotics. *arXiv* preprint arXiv:1701.00841. https://arxiv.org/pdf/1701.00841.pdf

Sedenberg, E., Chuang, J., & Mulligan, D. (2016). Designing commercial therapeutic robots for privacy preserving systems and ethical research practices within the home. *International Journal of Social Robotics, 8*(4), 575–587. https://doi.org/10.1007/s12369-016-0362-y

Smith, H. J., Dinev, T., & Xu, H. (2011). Information privacy research: An interdisciplinary review. *MIS Quarterly, 35*(4), 989–1016. https://doi.org/10.2307/41409970

Solove, D. J. (2008). Understanding privacy. Harvard University Press.

Stutzman, F., Capra, R., & Thompson, J. (2011). Factors mediating disclosure in social network sites. *Computers in Human Behavior, 27*(1), 590–598. https://doi.org/10.1016/j.chb.2010.10.017

Syrdal, D. S., Walters, M. L., Otero, N., Koay, K. L., & Dautenhahn, K. (2007). "He knows when you are sleeping"—Privacy and the personal robot companion. In *Proceedings of the 2007 AAAI Workshop Human Implications of Human–Robot Interaction,* Washington DC, 9–11 March 2007, pp. 28–33. AAAI. https://www.aaai.org/Papers/Workshops/2007/WS-07-07/WS07-07-006.pdf

Tamò-Larrieux, A. (2018). *Designing for Privacy and Its Legal Framework*. Springer.

Turkle, S. (2011). Authenticity in the age of digital companions. In M. Anderson & S. L. Anderson (Eds.), *Machine Ethics* (pp. 62–76). Cambridge University Press.

Van den Berg, B. (2016). Mind the air gap. In S. Gutwirth, R. Leenes, & P. De Hert (Eds.), *Data Protection on the Move: Current Developments in ICT and Privacy/Data Protection* (pp. 1–24). Springer.

Venkatesh, V., & Morris, M. G. (2000). Why don't men ever stop to ask for directions? Gender, social influence, and their role in technology acceptance and usage behavior. *MIS Quarterly, 24*(1), 115–139. https://doi.org/10.2307/3250981

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly, 27*(3), 425–447. https://doi.org/10.2307/30036540

Waldman, A. E. (2018). *Privacy as trust: Information privacy for an information age*. Cambridge University Press.

Warren, S. D. & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review, 4(*5), 193–220. https://doi.org/10.2307/1321160

Weiss, A., Wurhofer, D., & Tscheligi, M. (2009). "I love this dog"—children's emotional attachment to the robotic dog AIBO. *International Journal of Social Robotics, 1*(3), 243–248. https://doi.org/10.1007/s12369-009-0024-4

Westin, A. (1967). *Privacy and Freedom*. Atheneum Press.

Young, A. L., & Quan-Haase, A. (2013). Privacy protection strategies on Facebook: The Internet privacy paradox revisited. *Information, Communication & Society, 16*(4), 479–500. https://doi.org/10.1080/1369118x.2013.777757

Yu, R., Hui, E., Lee, J., Poon, D., Ng, A., Sit, K., Ip, K., Yeung, F., Wong, M., Shibata, T., & Woo, J. (2015). Use of a therapeutic, socially assistive pet robot (PARO) in improving mood and stimulating social interaction and communication for people with dementia: Study protocol for a randomized controlled trial. *JMIR Research Protocols, 4*(2). https://doi.org/10.2196/resprot.4189

Appendix A
Pictures Shown To Precipitants

FIGURE A Picture of Social Robot Interacting With Teenager

https://web.archive.org/web/20200125184849/https://assets.newatlas.com/dims4/default/6ca1072/2147483647/strip/true/crop/1000x677+0+0/resize/1000x677!/quality/90/?url=https%3A%2F%2Fassets.newatlas.com%2Farchive%2Fnaonextgen-1.jpg

FIGURE B Picture of Social Robot Interacting With Woman

https://web.archive.org/web/20190705035114/https://images.theconversation.com/files/99788/original/image-20151027-4997-1oqg5sv.jpg?ixlib=rb-1.1.0&rect=868%2C800%2C4131%2C2005&q=45&auto=format&w=1356&h=668&fit=crop

FIGURE C Picture of Social Robot Interacting With Children

https://web.archive.org/web/20190704053145/https://secure.i.telegraph.co.uk/multimedia/archive/03512/pepper-1_3512887b.jpg

FIGURE D Picture of Social Robot Interacting With Senior

https://web.archive.org/web/20200125190432/https://www.knowablemagazine.org/sites/default/files/styles/750_y/public/articles/content/2017-10/Paro_Japan.jpg?itok=znt9Id_z

Appendix B
Questionnaire

Table A Questionnaire and Items Used

Trusting Beliefs (based on McKnight et al., 2002)	***Please tell us how much you agree or disagree with the following statements.***
	I believe that the robots act in my best interest.
	If I required help, robots would do their best to help me.
	Robots perform their role of offering personal services really well.
	Robots are truthful in their dealings with me.
	Robots would keep their commitments.
Social Influence (based on Venkatesh, Morris, et al., 2003)	***For the following statements, imagine you had a robot at home such as one of those shown in the pictures at the beginning of the survey. Please tell us how much you agree or disagree with the following statements.***
	People who influence my behavior think I should use such a robot.
	People who are important to me think that I should use such a robot.
	In general, my friends have supported or would support the use of such a robot.
Perceived Benefits (Eurobarometer 2012, 2015)	***Here is a list of things that could be done by robots. For each of them, please tell us using a scale from 0 to 10, how you would personally feel about it. On this scale, 0 means that you would feel totally uncomfortable and 10 means that you would feel totally comfortable with this situation. Use the slider to select the number.***
	Having a robot assist you at work. (functional)
	Having a robot do household chores. (functional)
	Having a robot assist children with their homework. (functional)
	Using a robot in school as a means of education. (functional and emotional)
	Having a robot provide services and companionship to elderly people. (emotional)
Overall Privacy Concerns (based on Malhotra et al., 2004)	***Please tell us how much you agree or disagree with the following statements.***
	Overall, I see a real threat to my privacy due to the robot.
	I fear that something unpleasant can happen to me due to the presence of the robot.
	I do not feel safe due to the presence of the robot.
	Overall, I find it risky to have such a robot.

Informational Privacy Concerns (Social and Institutional) (first three items adapted from Stutzman et al., 2011, and last four items newly developed and partly based on Malhotra et al., 2004)	***Please indicate your level of concern about the following potential privacy risks that arise when you share your personal information with a robot.***
	Other users engaging in identity theft through the robot. (social)
	Other users hacking into the robot. (social)
	Other users stalking me via the robot. (social)
	The robot manufacturer insufficiently protecting personal data. (institutional)
	The robot manufacturer tracking and analyzing personal data. (institutional)
	The robot manufacturer selling personal data to third parties. (institutional)
	The robot manufacturer sharing personal data with government agencies. (institutional)
Physical Privacy Concerns (self-developed)	***Please indicate your level of concern about the following potential privacy risks that arise when you have a robot at home.***
	The robot damaging or dirtying my personal belongings (e.g., furniture).
	The robot asking me personal questions.
	The robot snooping through my personal belongings (e.g., pictures).
	The robot entering areas that it should not access (e.g., bedroom).
	The robot using items that it should not use (e.g., bedclothes, pillows, personal hygiene products).
Scientific Interest (Eurobarometer, 2012)	***Please tell us whether you are very interested, moderately interested, or not at all interested in scientific discoveries and technological developments.***

Table note: We relied on 5-point Likert scales ranging from "strongly disagree" to "strongly agree" for all items, except for privacy concerns. Here, respondents could assess their concern on a 5-point scale ranging from "no concern at all" (1) to "very high concern" (5)."

Appendix C
Measurement Model

Table B Measurement Model

Construct	Item	Std. Loading	t-values	R^2	α	C.R.	AVE	Descriptive Statistics
Intention to use Robots (INT)	int1 int2 int3 int4	0.889 0.846 0.799 0.678	50.656*** 40.211*** 36.668*** 22.314***	0.791 0.715 0.639 0.460	0.91	0.88	0.65	Mean: 3.15 Median: 3.50 Std. deviation: 1.32
Trusting Beliefs (TRUST)	trust1 trust2 trust3 trust4 trust5	0.829 0.848 0.754 0.730 0.733	34.740*** 44.124*** 26.658*** 25.503*** 22.102***	0.687 0.719 0.568 0.533 0.537	0.89	0.89	0.61	Mean: 3.59 Median: 3.80 Std. deviation: 1.07
Social Influence (SOI)	soi1 soi2 soi3	0.658 0.632 0.821	16.272*** 15.091*** 26.657***	0.433 0.399 0.674	0.86	0.75	0.50	Mean: 2.75 Median: 3.00 Std. deviation: 1.12
Perceived Benefits (BEN)	ben1 ben2 ben3 ben4 ben5	0.747 0.761 0.808 0.749 0.812	26.874*** 29.570*** 37.284*** 26.767*** 36.409***	0.558 0.579 0.653 0.561 0.659	0.88	0.88	0.60	Mean: 7.00 Median: 6.19 Std. deviation: 3.13 (0–10 scale)
Overall Privacy Concerns (OVP)	ovp1 ovp2 ovp3 ovp4	0.736 0.790 0.882 0.936	27.445*** 29.301*** 56.349*** 79.940***	0.542 0.625 0.778 0.877	0.90	0.91	0.71	Mean: 2.61 Median: 2.25 Std. deviation: 1.19
Privacy Concerns: Social (SOP)	sop1 sop2 sop3	0.827 0.796 0.722	29.542*** 23.428*** 20.244***	0.685 0.634 0.521	0.82	0.83	0.61	Mean: 3.17 Median: 3.33 Std. deviation: 1.22
Privacy Concerns: Institutional (INP)	inp1 inp2 inp3 inp4	0.820 0.904 0.905 0.830	34.186*** 59.371*** 63.209*** 31.180***	0.672 0.818 0.820 0.688	0.92	0.92	0.75	Mean: 3.72 Median: 4.00 Std. deviation: 1.18
Physical Privacy Concerns (PHP)	php1 php2 php3 php4 php5	0.565 0.743 0.883 0.893 0.814	13.276*** 22.010*** 44.099*** 54.327*** 31.291***	0.319 0.552 0.779 0.797 0.672	0.88	0.89	0.62	Mean: 2.22 Median: 2.00 Std. deviation: 1.17
Criterion		≥ 0.5	min*	≥ 0.4, < 0.9	≥ 0.7	≥ 0.6	≥ 0.5	

α = Cronbach's Alpha; C.R. = composite reliability; AVE = average variance extracted.
Average, median, and standard deviation calculated per item and then averaged across items for each construct; N=374.

Table C Discriminant Validity Test (Fornell Larcker Criterion)

	AVE	INT	TRUST	SOI	BEN	OVP	SOP	INP	PHP
INT	0.65								
TRUST	0.61	0.43							
SOI	0.50	0.36	0.38						
BEN	0.60	0.61	0.55	0.26					
OVP	0.71	0.24	0.25	0.10	0.39				
SOP	0.61	0.01	0.01	0.01	0.02	0.21*			
INP	0.75	0.02	0.00	0.03	0.01	0.15*	0.39*		
PHP	0.62	0.01	0.04	0.00	0.04	0.24*	0.27*	0.11*	
BOP	0.60	0.06	0.01	0.05	0.01	0.02*	0.01*	0.00*	0.05*

Table note: Squared correlations between the constructs are shown; AVE = average variance extracted; * = not used in the same model; correlations between INT, TRUST, SOI, and BEN computed in the OVP model

HUMAN-MACHINE

COMMUNICATION

HUMAN-MACHINE COMMUNICATION

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.7

Interlocutors and Interactions: Examining the Interactions Between Students With Complex Communication Needs, Teachers, and Eye-Gaze Technology

Rhonda McEwen[1], Asiya Atcha[1], Michelle Lui[1], Roula Shimaly[1], Amrita Maharaj[2], Syed Ali[1], and Stacie Carroll[3]

1 Institute of Communication, Culture, Information and Technology, University of Toronto, Mississauga, Mississauga, Ontario, Canada
2 Faculty of Information, University of Toronto, Toronto, Canada
3 Beverley Junior Public School, Toronto District School Board, Toronto, Canada

Abstract

This study analyzes the role of the machine as a communicative partner for children with complex communication needs as they use eye-tracking technology to communicate. We ask: to what extent do eye-tracking devices serve as functional communications systems for children with complex communication needs? We followed 12 children with profound physical disabilities in a special education classroom over 3 months. An eye-tracking system was used to collect data from software that assisted the children in facial recognition, task identification, and vocabulary building. Results show that eye gaze served as a functional communication system for the majority of the children. We found voice affect to be a strong determinant of communicative success between students and both of their communicative partners: the teachers (humans) and the technologies (machines).

Keywords: human-machine communication, interpersonal communication, eye-gaze communication, communicative partner, communication disability, special education, educational technology

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

Introduction

Advances in digital technologies provide opportunities for users to directly interact with software and devices, supporting human cognitive processes related to communication. This is potentially beneficial for users with cognitive deficits and/or physical disabilities. The case for human-machine communication is strong for people for whom the machine is not only a tool, but is an integral part of their expression and access to information. As examples, screen-readers have facilitated internet access for blind and vision-impaired users (Chandrashekar & Hockema, 2009); iPod touch and tablet devices have provided a means of expression for nonverbal children with autism (McEwen, 2014; Hourcade et al., 2012); location-based applications can support life-skills curriculum (e.g., attention, motivation) for students with developmental disabilities (Demmans Epp et al., 2015), and e-gaze glasses can support communicative interactions between blind and sighted people (Qui et al., 2016; Qui et al., 2018).

Yet even as technology designers continue to learn and find optimal approaches to meet the needs of a wide range of users, those with more complex disabilities remain hard to support as so much remains unknown about how learning occurs. Prior to the 1950s, people with communication deficits, particularly those classified as nonverbal, were not accommodated in formal education systems and were considered to be brain damaged and of lower intellect (Botting, 2004) best served by institutionalization. Pedagogical techniques were, and to some extent still are, based on oral and written skill delivery and demonstration. Without functional language, education was difficult and often abandoned for this population. For neuro-typical children, speech development occurs between the ages of 18 months to 3 years, and while it is a complex social process, speech development is part of the anticipated developmental stage of early childhood, with significant delays signaling potential physiological and/or neurological concerns (Sladen, 1974). Depending on the individual's capabilities, when speech is delayed, underdeveloped, or absent, other communication systems are called upon as substitutes, such as sign language or picture exchange communication. However, when physical disabilities are also factors, such as an inability to use the hands or control facial expressions, the ability to communicate is considerably more difficult. It is only within the past 80 years that cognitive science research on nonverbal communication provided indications that there are other mechanisms available for expressive and receptive communication for those with complex communication needs.

Eye-tracking devices with voice output have recently emerged as potentially useful assistive communication technologies for those who are nonverbal and unable to use their hands for command input. Despite a need for more research on alternate and technology-centered communication systems, there are few studies (Gilroy et al., 2017) about elementary school-aged children who have complex communication needs. This is due to several factors: the smaller number of research participants within public elementary school settings; more onerous human ethics protocols for researching this population; and the challenging nature of designing research of nonverbal and communicatively challenged children, where traditional research methodologies like interviews and observation are not easily implementable. Therefore, it requires an approach to data collection that involves a careful development of measures in environments familiar to the students.

When at school, children with complex communication needs have additional adult support in their classrooms in the form of teachers and educational assistants. These adults work closely with their students and become important communicative partners, especially in the situation where the adult to child classroom ratios are small. When communication technologies are present in this type of scenario, the classroom environment includes technical and non-technical elements that, in combination, present a fertile ground for research on communication. Literature on disabilities note that restrictions in participation in the venues available to others is the everyday experience of persons with disabilities and that communication media can play a part in reducing the barriers to participation (Ellis & Goggin, 2015). Studies in school settings show that educational technologies are being incorporated into the classroom with the goal to improve learning outcomes, particularly for learners with needs requiring alternate approaches (Demmans Epp et al., 2015; Edyburn, 2013; Goggin & Newell, 2005; McEwen, 2014). However, while a focus on educational technology can support better technology design and curriculum integration, studying the role that the technology itself plays in interaction is an understudied aspect and relevant to studies of human-machine communication. The latter is the focus of this paper, and the school setting does not suggest a focus on education, but is strategic as it provides access to an understudied population that aggregates in few other spaces.

Theoretical Framework

To frame this as a communication interaction study, we turn from the educational technology literature to draw from theories in Science and Technology Studies (STS) and Human-Machine Communication. Twentieth-century, Western-scientific traditions adopted a non-technical versus technical dichotomy as a foundational premise in academia (Grint & Woolgar, 1997; Suchman, 2008), and in so doing drew a boundary between the technical and the social-psychological. On one side of this binary are technological artifacts and, on the other, social entities—in other words machines versus humans. The epistemologies that supported this construct included technological determinism on one side and humanist perspectives on the other. Each theoretical approach struggles to reposition either technology or people in the center of the analysis. Personal digital media that are deeply embedded in daily communication have called this conceptual separation into question. In the works of Vygotsky (1978) and later Latour and Woolgar (1979), STS scholars consider meaning-making as occurring within a particular social context. For scholars of this tradition the social context, which includes all of the elements in the communicative environment, is the focus from which technological and human interactions may be understood. The familiarity from everyday use of digital devices obscures their role in communication to human participants and observers. A key aspect of the theoretical framing of this study is the notion that when we use a digital technology for communication, we are also engaged in communication with the device itself.

Extending Niklas Luhmann's (1992) definition of communication that considers the bidirectional understanding that must occur for successful communication, we consider the elements of human-machine communication that occur when we engage in mediated communication. These elements include the affordances of the technologies, and the abilities of the users (Dubé & McEwen, 2016). Along with scholars who similarly posit that sociotechnical interactions are co-constituted, Wanda Orlikowski (2007) believes that neither humans nor technologies should be privileged in research analyses. Following from works of Suchman (2008) and Barad (2007), Orlikowski (2007) claims that in constitutional entanglement, ". . . the social and the material are considered to be inextricably related—there is no social that is not also material, and no material that is not also social" (p. 1437). Likewise, for human-machine communication scholars the distinction of humans and technologies is purely an abstraction since these entities relationally enact each other in everyday practice (Guzman, 2018). Drawing from this co-constitutive and human-machine communication theoretical frame, we designed our study of the interpersonal communication between teachers and students to also take into account the eye-tracking technology itself. The technology is not considered as simply a mediating device, but an active participant in the communication taking place.

Background

For the purposes of our study, we define the three communication units involved as illustrated below: (1) Individuals with complex communication disabilities, (2) the human communicative partner, including teachers, educational assistants, and therapists, and (3) the machine or assistive technology that enables the communication and supports the interactions for individuals with complex communication disabilities.

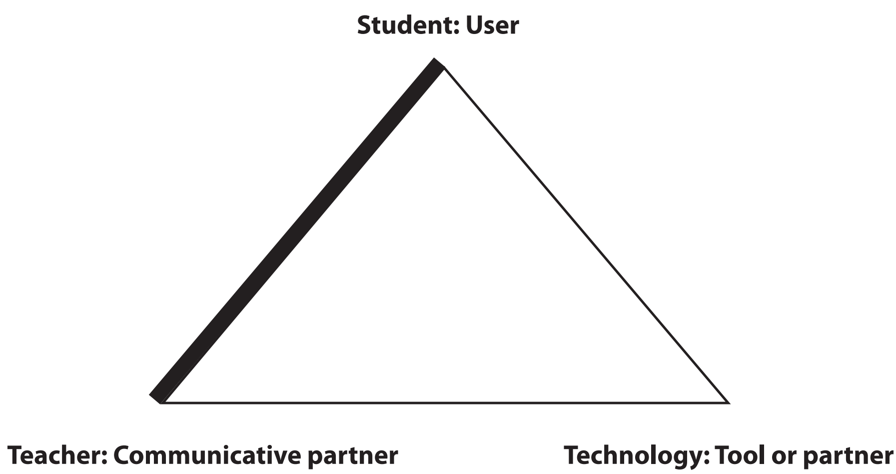


FIGURE 1 Communication Units

I. Student: Users With Complex Communication Disabilities

The first unit of analysis is the user and, in this case, students with complex communication disabilities. Children with multiple disabilities such as language, motor, and other impairments do not develop speech skills as expected and have limited opportunities for communication (Light, 1997). We follow 12 students with complex communication disabilities who use augmented and alternative communication (AAC); that is, communication techniques used to supplement or substitute spoken or written communication for those with impairments. Four out of the 12 students are officially diagnosed with Rett Syndrome (RS), and others with cerebral palsy, chromosome deletion, seizure disorder.

Eye gaze was reported as the most commonly used modality for expressive communication for individuals with RS (Bartolotta et al., 2011; Urbanowicz et al., 2014), which raises the opportunity to explore eye-tracking technology as a means to enable communication and improve communicative capabilities for individuals with RS. In a study of nonverbal cues, eye gaze was identified as a key feature in following conversational sources including in mediated exchanges (Vertegaal, 1999) and in an analysis of conversational attention in multiparty conversations, eye gaze was found to be an excellent predictor of conversational attention (Vertegaal et al., 2001). Another study explored the application of eye-tracking technology to analyze the intentionality of gaze of seven girls diagnosed with RS. Results show that eye gaze was intentionally used to perform three cognitive tasks with high accuracy, suggesting that eye gaze could be used for communication by people with similar limitations (Baptista et al., 2006).

Based on the existing literature, and given a gap in the literature since more recent evolution of eye-tracking technologies, our first research question (RQ1) asks: To what extent do eye-tracking devices serve as functional assistive communication systems for students with complex communication needs?

II. Teacher: The Communicative Partner (Human)

The second unit of analysis is the person/human with whom the student is communicating with; in this case, the teacher or educational assistant. For the purposes of simplicity, we will use the term teacher throughout this article to describe the adult educator working with the student. Mackenzie and Stoljar (2000) point out that "persons are socially embedded and that agents' identities are formed within the context of social relationships and shaped by a complex of intersecting social determinants . . ." (p. 4). Within school environments teachers can be considered to be important agents whose identities are partially shaped by interaction with each other and in day-to-day encounters with their students (Fredricks et al., 2004).

Communicative Partner's Language Style

Conclusions from prior research indicate that girls with RS learned to communicate more frequently and intentionally as a result of storybook reading with their mother (Skotko et al., 2004). The mothers were trained to attribute meaning to the girls' attempts to communicate, ask communicative questions, and prompt the use of communication devices or symbols through natural questions and comments rather than commands. This style of communication is more naturalistic and the use of an inflected tone, one with excitement and higher than usual emotional content, is more engaging. When parents used this inflected voice approach during storytelling it resulted in an increased number of communication attempts and communication means.

However, this approach is not standardized and is sometimes actively discouraged in the developmental communication literature. In studies of children with cognitive impairments researchers found that neutral voice affect mitigated difficulty that some children have with identifying the appropriate prosody or emotion of the speaker's words (Hobson et al., 1989; Stewart et al., 2013). Another study linked a specific region in the brain as the site where processing of prosody appears to be negatively impacted for people with neurological impairment, with recommendations for the use of neutral and uninflected tones to reduce the cognitive load (Wang et al., 2007).

Guided by this debate in the literature our second research question (RQ2) asks: What role does the communicative partner's language style (voice affect inflected vs. neutral) play in communicative outcomes when using eye-gaze technology?

Communicative Partner Familiarity

In a survey distributed to 141 parents, teachers, and health care professionals, the majority of respondents believed that people familiar to individuals with communication disabilities can better interpret their communication than unfamiliar people (Bartolotta et al., 2011). While familiarity offers more comfort in a social environment, in special education where there is greater likelihood that a nonverbal student's needs require some assistance and interpretation, there is a risk for the student to lose degrees of agency and self-determination. Facilitated communication, a method in which people who lack functional speech, usually due to a developmental disability, input commands into a device with the assistance of a facilitator (Stock, 2011; Wheeler et al., 1993), is a controversial issue. At the heart of the debate is the potential loss of independence and agency for the person being assisted—is their voice being heard or is it being directed by facilitators who are familiar with the person communicating? In addition, there are instances where the communicative partner may encounter challenges interpreting and identifying communication intentionality.

Identifying the intentionality of communication in individuals with autism or other communication disabilities is often exacting (Iacono et al., 1998) and communicative partners can exhibit inconsistencies in identifying behaviors that serve as a mean of communication (Mattews-Somerville & Cress, 2005). In addition, familiar partners are not always available necessitating communication systems that can be generalized to persons who may not know the individual trying to communicate.

Based on this, our third and final research question (RQ3) asks: What role does the familiarity (or unfamiliarity) of the communicative partners play in communicative outcomes when using eye-gaze technology?

III. Eye-Gaze Technologies

According to Goossens and Crain (1987), numerous eye-gaze communication techniques have appeared in literature since Eichler, McNaughton and Kates, and Vanderheiden (p. 77). The introduction of electronic eye-tracking systems allows the computer to handle almost the entire process of decoding the gaze for the purpose of message selection and confirmation. Because the eye-tracking software also acts as a speech-generating device (SGD), the child can initiate a conversation by gazing at an object on the screen, prompting computer-generated speech. The partner can focus on responding to the message that the child is communicating without having to also verbalize what they see the child is looking at (Gillespie-Smith & Fletcher-Watson, 2014). The child no longer needs to look at the human communicative partner to initiate a message or rely on them to determine the path of their visual attention. Gaze toward the human communicative partner becomes more socially weighted, such as communicating a sense of interest in the person themselves or to express excitement in the conversation (Djukic & McDermott, 2012). Another consequence of using technology for users with complex communication needs to increase their levels of *agency* which can be defined as "being in a state of action or exercising power, or as being free to choose and act in a manner independent of the structures that limit or influence the opportunities that individuals have" (García Carrasco et al., 2015, p. 162). Given the one-on-one interaction between the participant and the eye-gaze technology, research question 1 focuses on the participant-device interaction.

Method

The data were collected in 2016–2017 in a special needs school in downtown Toronto, Canada. The Tobii-Dynavox eye-tracker system was used to collect data from software designed to assist the children in facial recognition, scanning, targeting, and task identification. Our data collection used The Tobii-Dynavox eye-tracker system which had two configurations: (a) an embedded system, the Tobii I-12 that came from the manufacturer on a stand which could be adjusted for height and angle and had a camera built into the screen; and (b) an improvised hardware system, which included a myGaze camera, Tobii-Dynavox software, and a laptop mounted to a portable stand.

FIGURE 2 Embedded System, Participant and Communicative Partner Example

The 12 participants used an educational software designed to improve proficiency in communication using images with associated labels or phrases. For example, a photo or line drawing of a dog would have the label "dog" typed below it and a pre-recorded, audible output of the word "dog" from the device if that image is selected by the user using eye gaze. Similarly, in a graphic image with many objects, eye gaze detected by the device on the image would result in the auditory output of the label or phrase pre-recorded for the object. Data were collected in familiar surroundings during scheduled instructional times in three classrooms and during regular school hours.

Participants

Participants were nonverbal students with a range of cognitive or developmental disabilities *(see Table 1), were aged 4 to 12, and met the following inclusion criteria: (1) had difficulty using their hands (i.e., they could not easily input commands into a device), (2) had limited speech and had little to no spoken language ability, (3) were sighted (i.e., have functional vision, where prescription glasses were allowed).

TABLE 1 Participant characteristics and diagnosis (n = 12)

Participant Code	Classroom ID	Gender	Age	Diagnosis
C04	112	F	4	Rett syndrome
L05	112	F	5	Rett syndrome
T06	112	M	6	Complex, not otherwise specified
Z07	112	F	7	Rett syndrome
A08	113	F	8	Complex, not otherwise specified
K05	113	F	5	Rett syndrome
N05	113	F	5	Cerebral palsy
R07	113	M	7	Chromosome deletion q13
R08	113	F	8	Brain injury
A11	116	F	11	Cerebral palsy
E09	116	F	9	Brain injury
L12	116	F	12	Seizure disorder

Prior to this study students had access to both low- and high-tech AAC devices in the classroom. Most participants (n = 11) had experience with some modes of analog and digital eye-gaze tracking communication. In the analog systems, students engaged in the selection of pictures with the aid of a communicative partner. For example, the communicative partner would hold two objects in front of the child and the child would direct their gaze toward their choice. The students also had some practice in establishing joint attention.

Procedure

The participants' social and communication skills were baselined using the Communication Matrix—an online communication assessment created for emergent communicators and those who use alternative communication systems (Rowland, 2011, https://www.communicationmatrix.org/). The Communication Matrix is a detailed assessment tool with categories defined for the skill identification of pre-verbal communicators and has been successfully used by the research team in previous studies (McEwen, 2014). Pre- and post-assessments were completed to track changes in communication development over the course of the project. This form of assessment is useful when chronological or developmental age normative classification are ambiguous and/or misleading. When studying children with disabilities using intergroup profiles of normative skill acquisition is not applicable since the chronological ages and developmental ages do not often match (Rutter, 1989; Tsao &

Kindelberger, 2009). Instead, we consider inter-individual variability in their cognitive functioning by using tools like the Communication Matrix in pre-post design.

Before the research project started ethics clearance was received from the Toronto District School Board and from the University of Toronto, including written parental consent as part of the protocol. Data were collected using video and screenshots that were loaded onto an assessment tool for educators (SesameSnap) and stored in a password protected online format. Data were also collected in the Communication Matrix software where the system assigned them a random ID, which the research team appended with the assigned anonymized ID codes as previously described.

Analytical Measures

Four measures were selected to investigate the research questions. RQ1 is concerned with the extent to which eye-tracking devices can be used for functional assistive communication. *Session time* is defined as the total amount of time that the student participant and device are engaged in a communicative interaction, measured in minutes and seconds. Several previous studies (DeVito & DeVito, 2007; Duck et al., 1988; Emmers-Sommer, 2004; Luhmann, 1992) found that prolonged communication is a significant predictor of successful interactions. Therefore, longer session times would indicate interest, motivation, and overall effectiveness in communication, especially between the device and the participant. RQ2 asks what role the communicative partner's language style or voice affect plays in communicative outcomes when using eye-gaze technology. As discussed in the earlier background section, previous research findings are contradictory regarding neutral versus inflected tones of voice used by the communicative partner, thus we examine interactions with *inflected* and *neutral* tones of voice. We trained communicative partners in the use of high and low affect voice, using voice samples to maintain consistency. Finally, RQ3 considers how *familiarity* between the communicative partner and the participant affects communicative outcomes, thus whether or not the communicative partner is familiar to the student is the final variable under investigation. This is measured in the amount of contact time that the communicative partner had with the participants prior to the study, with minimal contact (less than 1 hour per week), average contact (between 1 and 3 hours per week), and high contact (over 3 hours per week) as the classification points.

Data Collection

Data collection was conducted by the three classroom teachers over a 3-month period. Teachers videotaped the students while they used the eye-tracking devices. Data collection sessions occurred twice a week. Participants were tasked with using software with phrases and labels to identify objects. One application is called Sono Primo software (Tobii Dynavox Ltd.) which includes eye-tracking software for developing AAC skills. Learners look around interactive scenes (e.g., farm, birthday party) and the visual targets play related sounds, including phrases and labels, when triggered. Familiar partners were the students' regular teachers, who spent time with them during regular classroom instruction. Unfamiliar partners were other teachers or assistants who worked elsewhere in the school. Unfamiliar partners were familiar with the instructional environment and have experience working with children with disabilities and developmental delays, but did not know the students personally. Both partner types used a mix of inflected and neutral tones in different sessions, to allow for researchers to analyze the impact of both variables against the different partner types.

Coding and Analysis

All of the video data were coded independently by three final-year undergraduate students supervised by the principal researcher. Three undergraduate researchers from the team started by selecting a sample of videos from the three classrooms for comparative assessment as a group to synchronize the coding for inter-rater reliability. They made qualitative notes on the videos, looking for variables that could be used to assess the validity of the hypotheses that followed from the research questions. A codebook was developed to guide the rest of the coding process. The team met twice for calibration and inter-coder reliability checks. The data were analyzed using IBM SPSS software, using a paired-samples t-test for comparing pre- and post-intervention Communication Matrix assessments.

Linear Generalized Estimating Equations (GEE) analyses, which is appropriate for the analysis of data collected in repeated measures designs (Ballinger, 2004), were used to determine whether the duration of eye-gaze sessions is predicted by voice affect and partner type. Seven videos from four of the participants (K05, C04, A08, and E09) were transcribed to explore the quality of interactions within the sessions. Using a multimodal interaction methodological framework (Norris, 2004), the transcriptions include descriptions of the surrounding environment, the position of the participants and their partners, nonverbal utterances and facial expressions, and body movements. Multimodal analysis describes the use of data from gestures and movement in communication—this is an important consideration in all communication but more so in communication on nonverbal people.

Results

Communication Matrix assessments: RQ1 questions the extent to which the eye-gaze system in use served as a functional assistive communication device for students with complex communication needs. Eleven of 12 students completed both pre- and post-intervention Communication Matrix assessments. Figure 3 below presents comparative pre- and post-intervention Communication Matrix scores, which shows that 8 of 11 students who used the eye-gaze technology showed improvement in their communication skills, while one showed no change. According to a paired samples t-test, we found a significant difference in the scores for the pre-test ($M = 41.00$, $SD = 23.99$) and post-test ($M = 52.91$, $SD = 34.90$), $t(10) = 3.01$, $p < .05$.

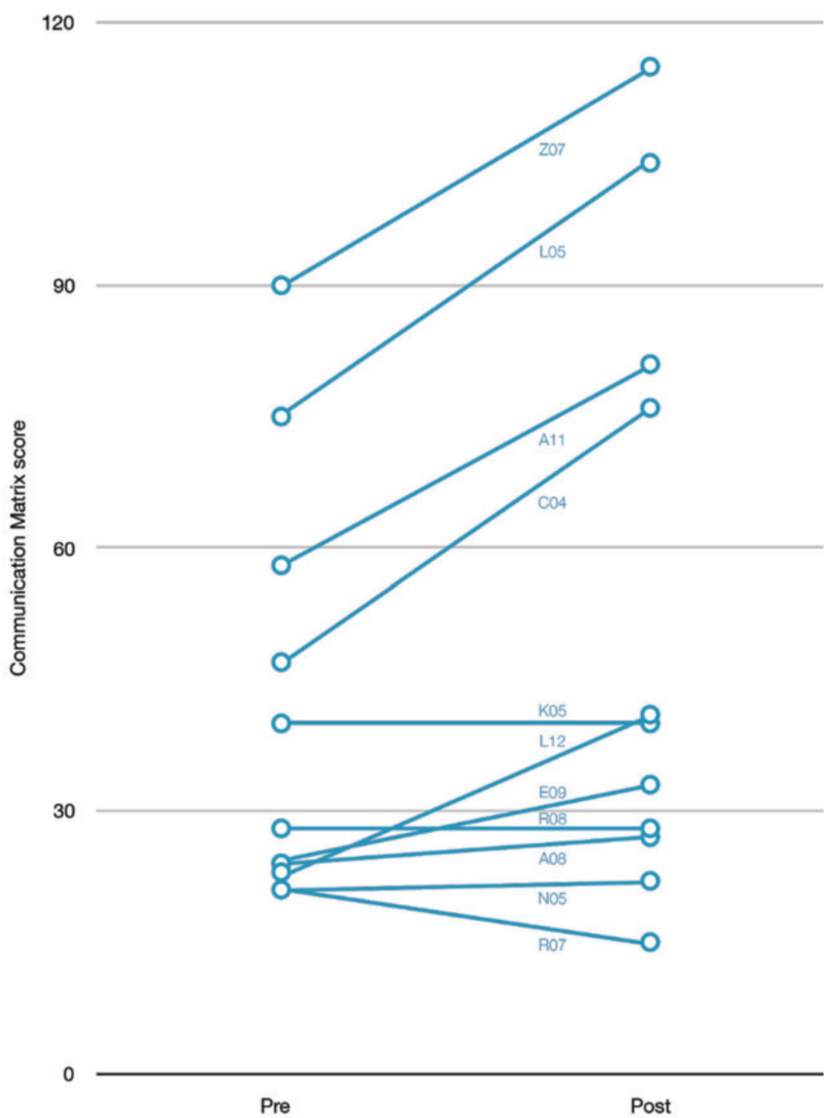


FIGURE 3 **Pre- and Post-Intervention Communication Matrix Scores (n = 11)**

Session Time

Regarding RQ1, capturing variable session length produced valuable results. A total of 164 testing sessions were conducted with an average number of 9.25 sessions per student (SD = 7.35; range = 2–24), lasting an average of 154.54 seconds (SD = 32.20). Teachers confirmed that these interactions were longer than is typical, where on average non-eye gaze mediated interactions are between 35–50 seconds. When cross-tabulating session time with the variable of voice affect from the second research question (RQ2), we found that voice affect significantly influenced the length of interactions (summarized in Table 2).

Table 2 Descriptive Statistics for Session Time Data (seconds)

	Number of Sessions	Mean Session Time (seconds)	Standard Deviation
All sessions	164	154.54	32.20
Communicative Partner Type			
Familiar	60	134.47	106.61
Unfamiliar	29	111.76	87.92
Voice Affect*			
Inflected	59	148.61	105.25
Neutral	25	92.96	44.87

Table 3 Mean Session Time (seconds)
Data Organized by Communicative Partner and Voice Affect Variables

	Communicative Partner Type								
Voice Affect	Familiar			Unfamiliar			Both		
	M	SD	n	M	SD	N	M	SD	n
Inflected	150.28	95.58	43	144.13	131.25	16	148.61	105.25	59
Neutral	95.94	47.51	16	87.67	41.94	9	92.96	44.87	25
Both	134.47	87.92	59	111.76	106.61	25	n/a		

General Estimating Equations were used to evaluate the effect of voice affect on session time, which found that sessions with infected tone were predicted to be significantly longer than sessions when communication partners used a neutral voice ($p < .001$). When cross-tabulating session time with the variable of familiarity of the communicative partner from RQ3, we found that sessions conducted by a familiar communicative partner were not significantly longer ($M = 134.47$; $SD = 106.61$) than those conducted by an unfamiliar communicative partner ($M = 111.76$; $SD = 87.92$), as determined by the GEE approach ($p = .307$).

Further analysis (summarized in Table 3) indicates that sessions with a familiar communicative partner using an inflected tone were associated with the longest sessions ($M = 150.28$, $SD = 95.58$), followed by sessions with an unfamiliar partner using an inflected voice ($M = 144.13$, $SD = 131.25$). Sessions in which communicative partners used a neutral tone were shorter, with familiar partners sessions being slightly longer ($M = 95.94$, $SD = 47.51$) than unfamiliar partners on average ($M = 87.67$, $SD = 41.94$).

Multimodal Analysis

Results of the Multimodal Analysis also respond to RQ1. At seven minutes and eleven seconds into the session the following example comes from one of the longest interactions with a child and a familiar partner.

Teacher: What else is up here? [short pause]

Eye-gaze system: Can you help me?

Teacher: Help you do what? We just did brush your teeth. What else should we do?

[Camera pans back toward C04, C04 was looking at the teacher but turns head slightly back toward system]

Eye-gaze system: [*cursor lights up around the picture of a faucet*] I need water.

Teacher: (pretends to gasp) Oh. We can get your actual drink if you want a drink. Yeah, I'm happy to get you a drink.

In this example teacher is very expressive and uses a highly inflected voice throughout the interaction. The recording starts with the teacher repeating a request made by the child for water and pretending to pour some for the student. The child responds with a verbalization and the partner prompts her to try a different request. We hear the software in the background say, "Can you help me?" This lights up a stick figure with a toothbrush and the teacher offers to help the child with brushing her teeth. She mimes it with animated sounds, making the child smile. We see that the child will repeat the request for water through the eye-tracking system and the teacher will suggest getting a real drink in case she is thirsty.

We note that the immediate responses by the teacher to the communication by the student reduces the time between prompt and response. There were several instances with other participants where this became evident, possibly associating voice affect with reduced time between prompts and response with evidence of prompts fading over time; however, this requires further investigation.

Discussion

From the results we can annotate the initially proposed model of eye-gaze communication (see Figure 4). To answer RQ1, the data show that eye-gaze communication was a functional assistive communication system for the majority of the students with complex communication needs. Results show that students with complex communication needs are able to engage in richer exchanges with the device and teachers.

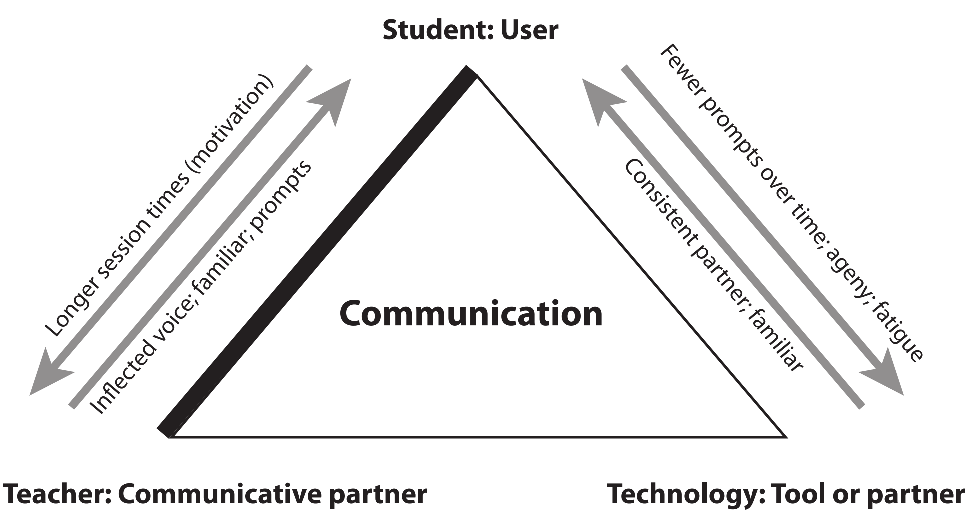


Figure 4 Modified Model of Eye-Gaze Communication

Communication Matrix data showed communication skill gains for eight out of 11 students; with one student staying the same. It is not clear why two students showed skill reductions, but since they were small changes it is possible that this indicates an error in measurement.

Regarding RQ2, we found voice affect to be a strong determinant of the interactions between students, their communicative partner, and the technology. While there is debate in the literature on whether or not neutral or inflected voice styles are recommended, the data in this research show that inflected responses by communicative partners were significantly correlated with longer engagements. When teachers used more inflected tones, students were also more expressive, visibly enthusiastic, and neither seem to be confused nor frustrated with the additional information within the exchanges. This suggests that with respect to eye-gaze technology use by students with diverse and complex communication needs, the communication style of the teacher plays a key role in positive outcomes.

The final research question, RQ3, focused on the role that the familiarity of the teacher played in communication. Results show that although sessions were, on average, slightly longer with familiar communicative partners than with unfamiliar partners, the difference was not significant. This is a somewhat surprising result as we expected a greater qualitative difference between sessions with familiar partners and unfamiliar partners, based on prior literature on familiar caregivers' role in interpreting communicative acts of children with communication disabilities (e.g., Sigafoos et al., 2011). Yet, our findings support prior studies that examined familiar and unfamiliar adults' interpretations of "potential" communicative acts in learners with RS (Julien et al., 2015) and found that the majority of familiar and unfamiliar adults were able to recognize potential communicative behaviors. This is an important finding because it suggests that among teachers in special education it matters less that students know them personally than it does the way that they engage with students. While we do not suggest that the quality of interactions between familiar and unfamiliar partners are comparable, we note that any issues encountered with unfamiliar partners were not problematic enough to cause breakdown of communication. One possible explanation is that although unfamiliar communicative partners were unfamiliar to the learners, they were familiar with the school and were well trained to address the communicative needs of the students in general.

Throughout the research project students demonstrated increased joint attention and reciprocal communication through eye-gaze with an expressive communicative partner through longer session times. In a form of cause and effect, students learned that they could make the system respond to their eye-gaze in a consistent manner that became familiar to them over time. For some students who have very little control over their bodies and, indeed, many aspects of their lives they could control a technology. The video data clearly showed an increase in agency by the students and an empowerment that likely contributed to their motivation to communicate, even when fatigue was also an outcome. It is possible that teachers would need to be mindful of students becoming overly reliant on the eye-gaze technology in the future; however, this study indicates that eye-gaze technologies have a role to play in the repertoire of assistive communication devices for children with severe deficits.

Applying a triad analysis to this research allowed us to pay more attention to the role that the technology played in the communicative exchanges. The eye-gaze technology was not simply a tool but was also a communicative partner to the student. The student was engaged in an exchange with the technology that required both sides of that dyad to have an accurate understanding of what was necessary for success. The students learned that by looking at the screen and holding their gaze, a voice output resulted that highly engaged their teachers.

When the technology worked according to plan, its affordances of consistency and reliability made it a familiar communicative partner and offloaded some of the effort on the part of the teacher in the exchanges. The technology as an active communicative partner appears to have changed the dynamic between the student and teacher, in some cases mitigating losses due to unfamiliarity between student and teacher. The technology became the consistent factor in the communicative system. For the student there were two communicative partners. This may be a factor in the level of fatigue that they displayed at the end of the sessions. Further study is needed to understand this outcome.

Conclusion

This research and its findings contribute to a small but growing literature on communication for people with complex needs. However, there were a number of limitations within the project that should be noted. There is some inconsistency in the number of sessions conducted within each set of variables. The staff's ability to collect data was subject to resource availability and was scheduled to minimize impact on classroom routines. There are more sessions with familiar partners, for instance, because they were the homeroom teachers and consequently available more often.

This is a preliminary study, and the sample size is small, albeit typical of a school setting that supports learners with this cluster of rare disorders. That said, we believe our findings may be more generalizable to other educational settings and reflect more genuine communicative contexts than other studies conducted in lab conditions with participants from the more general population. The results would benefit from replicated studies of RS learners using eye-tracking devices in other educational settings.

Interaction with complex, naturalistic scenes can be considered a form of gameplay, which is a high-energy interaction and demands quite a lot from the participants, particularly when the communicative partner uses an inflected voice. The resulting peak in communication skills could be short-lived, and children may return to normal modes of communication in day-to-day activities, with significantly less engagement through the communicative partner's voice affect in other contexts. It encourages them to play and to interact while in game mode, but may not be suitable in all interactions. That said, teachers did notice significant improvements in communication outcomes six months after the testing was concluded. In future studies, it may be helpful to substantiate this by completing a communication matrix for the participants 6 months out from the initial testing period. As technologies such as eye-tracking devices emerge it is important to identify the factors that affect outcomes. Studies that attend to young learners with more severe communication deficits can lead to implementable solutions at the school level with immediate positive impacts. The criticality of the role of the communicative partner is repurposed with the assistance of eye-gaze technology. The technology becomes an active agent, more than a passive tool, for the teachers to enable deeper engagement with students with complex communication needs.

Acknowlegments

This project was funded by the Social Science and Humanities Research Council of Canada (SSHRC), under grant #950-231395.

Please address correspondences to: Prof. Rhonda McEwen, rhonda.mcewen@utoronto.ca, University of Toronto Mississauga, Institute of Communication, Culture, Information and Technology, 3359 Mississauga Road, Mississauga, ON L5L 1C6

There was no potential conflict of interest that affect the research reported in the enclosed paper.

Author Biography

The primary investigator, **Dr. Rhonda McEwen**, is Canada Research Chair in Tactile Interfaces, Communication and Cognition and is an Associate Professor of New Media & Communication. Dr. McEwen is the Director of the Institute of Communication, Culture, Information and Technology at the University of Toronto Mississauga. With an MBA in IT from City University in London, England; an MSc in Telecommunications from the University of Colorado; and a PhD in Information from University of Toronto, Dr. McEwen combines communication studies, applied, and behavioral sciences to examine the social and cognitive effects of technologies. Her pioneering approach to using experimental techniques, eye tracking, observations, sensor data, and interviews to investigate Human-Machine Communication involving children and adults across the user spectrum, and including those diagnosed with communication and learning disorders.

https://orcid.org/0000-0002-9451-2909

References

Ballinger, G. A. (2004). Using generalized estimating equations for longitudinal data analysis. *Organizational Research Methods, 7*(2), 127–150. https://doi.org/10.1177/2F1094428104263672

Baptista P. M., Mercadante M. T., Macedo E. C., Schwartzman J. S. (2006). Cognitive performance in Rett syndrome girls: A pilot study using eye tracking technology. *Journal of Intellectual Disabilities Research*, *50*(9), 662–666. https://doi.org/10.1111/j.1365-2788.2006.00818.x

Barad, K. (2007). *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Duke University Press.

Bartolotta, T. E., Zipp, G. P., Simpkins, S. D., & Glazewski, B. (2011). Communication skills in girls with Rett Syndrome. *Focus on Autism and Other Developmental Disabilities*, *26*(1), 15–24. https://doi.org/10.1177/1088357610380042

Botting, N. (2004). Non-verbal cognitive development and language impairment. *Journal of Child Psychology and Psychiatry*, *46*(3), 317–326. https://doi.org/10.1111/j.1469-7610.2004.00355.x

Chandrashekar, S., & Hockema, S. (2009). Online access, participation and information credibility assessment. *Proceedings of the Science and Technology for Humanity (TIC-STH), IEEE.* Toronto International Conference, Toronto, 26–27, September 2009. https://doi.org/10.1109/TIC-STH.2009.5444397

Demmans Epp, C., McEwen, R., Campigotto, R., & Moffatt, K. (2015). Information practices and user interfaces: Student use of an iOS application in Special Education. *Education and Information Technologies*, *21*(5), 1–24. https://doi.org/10.1007/s10639-015-9392-6

DeVito, J. A., & DeVito, J. (2007). *The interpersonal communication book*. Pearson/Allyn and Bacon.

Djukic, A., & McDermott, M. V. (2012). Social preferences in Rett Syndrome. *Pediatric Neurology*, *46*(4), 240–242. https://doi.org/10.1016/j.pediatrneurol.2012.01.011

Dubé, A. K., & McEwen, R. N. (2017). Abilities and affordances: factors influencing successful child–tablet communication. *Educational Technology Research and Development*, *65*(4), 889–908. https://doi.org/10.1007/s11423-016-9493-y

Duck, S., Hay, D. F., Hobfoll, S. E., Ickes, W., & Montgomery, B. M. (Eds.). (1988). *Handbook of personal relationships: Theory, research and interventions*. John Wiley & Sons.

Edyburn, D. L. (2013). Critical issues in advancing the special education technology evidence base. *Exceptional Children*, *80*(1), 7–24. https://doi.org/10.1177/001440291308000107

Ellis, K., & Goggin, G. (2015). Disability media participation: Opportunities, obstacles and politics. *Media International Australia*, *154*(1), 78–88. https://doi.org/10.1177/1329878X1515400111

Emmers-Sommer, T. M. (2004). The effect of communication quality and quantity indicators on intimacy and relational satisfaction. *Journal of Social and Personal Relationships*, *21*(3), 399–411. https://doi.org/10.1177/0265407504042839

Fredricks, J., Blumenfeld, P., & Paris, A. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109. https://doi.org/10.3102/00346543074001059

García Carrasco, J., Hernández Serrano, M. J., & Martín García, A. V. (2015). Plasticity as a framing concept enabling transdisciplinary understanding and research in neuroscience and education. *Learning, Media and Technology*, *40*(2), 152–167. https://doi.org/10.1080/17439884.2014.908907

Gillespie-Smith, K., & Fletcher-Watson, S. (2014). Designing AAC systems for children with autism: Evidence from eye tracking research, *Augmentative and Alternative Communication*, *30*(2), 160–171. https://doi.org/10.3109/07434618.2014.905635

Gilroy, S. P., McCleery, J. P., & Leader, G. (2017). Systematic review of methods for teaching social and communicative behavior with high-tech augmentative and alternative communication modalities. *Review Journal of Autism and Developmental Disorders*, *4*(4), 307–320. https://doi.org/10.1007/s40489-017-0115-3

Goggin, G., & Newell, C. (2005). *Disability in Australia: Exposing a social apartheid*. UNSW Press.

Goossens, C. A., & Crain, S. S. (1987). Overview of nonelectronic eye-gaze communication techniques. *Augmentative and Alternative Communication*, *3*(2), 77–89. https://doi.org/10.1080/07434618712331274309

Grint, K., & Woolgar, S. (1997). *The machine at work: Technology, work and organization*. Polity Press.

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

Hobson, R. P., Ouston, J., & Lee, A. (1989). Naming emotion in faces and voices: Abilities and disabilities in autism and mental retardation. *British Journal of Developmental Psychology*, *7*(3), 237–250. https://doi.org/10.1111/j.2044-835X.1989.tb00803.x

Hourcade, J. P., Bullock-Rest, N. E., and Hansen, T. E. (2012). Multitouch tablet applications and activities to enhance the social skills of children with autism spectrum disorders. *Personal and Ubiquitous Computing, 16*(2), 157–168. https://doi.org/10.1007/s00779-011-0383-3

Iacono, T., Carter, M., & Hook, J. (1998). Identification of intentional communication in students with severe and multiple disabilities. *Augmentative and Alternative Communication*, *14*(2), 102–114. https://doi.org/10.1044/aac10.2.7

Julien, H. M., Parker-McGowan, Q., Byiers, B. J., & Reichle, J. (2015). Adult interpretations of communicative behavior in learners with Rett syndrome. *Journal of Developmental and Physical Disabilities*, *27*(2), 167–182. https://doi.org/10.1007/s10882-014-9407-z

Latour, B., & Woolgar, S. (1979). *Laboratory life: The social construction of scientific facts. Sage*.

Light, J. (1997). "Let's go star fishing": Reflections on the contexts of language learning for children who use aided AAC. *Augmentative and Alternative Communication*, *13*(3), 158–171. https://doi.org/10.1080/07434619712331277978

Luhmann, N. (1992). What is communication? *Communication Theory*, *2*(3), 251–259. https://doi.org/10.1111/j.1468-2885.1992.tb00042.x

Mackenzie, C., & Stoljar, N. (2000). *Relational autonomy: Feminist perspectives on autonomy, agency, and the social self*. Oxford University Press.

Mattews-Somerville, R. C., & Cress, C. J. (2005). Parent perceptions of communication behaviors at formally assessed stage transition in young children at risk for being non-speaking. *Communication Disorders Quarterly, 26*(3), 164–177. https://doi.org/10.1177/2F15257401050260030501

McEwen, R. (2014). Mediating sociality: The use of iPod Touch™ devices in the classrooms of students with autism in Canada. *Information, Communication & Society*, *17*, 1264–1279. https://doi.org/10.1080/1369118X.2014.920041

McEwen, R., Zbitnew, A., & Chatsick, J. (2016). Through the lens of a tetrad: Visual storytelling on tablets. *Educational Technology & Society*, *19*(1), 100–112.

Norris, S. (2004). *Analyzing multi-modal interaction: A methodological framework*. Routledge.

Orlikowski, W. J. (2007). Sociomaterial practices: Exploring technology at work. *Organization Studies*, *28*(9), 1435–1448. https://doi.org/10.1177/0170840607081138

Qiu, S., Anas, S. A., Osawa, H., Rauterberg, M., & Hu, J. (2016). E-gaze glasses: simulating natural gazes for blind people. In *Proceedings of the TEI'16: Tenth International Conference on Tangible, Embedded, and Embodied Interaction, ACM,* 563–569.

Qiu, S., Han, T., Rauterberg, M., & Hu, J. (2018). Impact of simulated gaze gestures on social interaction for people with visual impairments. In *Transdisciplinary Engineering Methods for Social Innovation of Industry 4.0: Proceedings of the 25th ISPE Inc. International Conference on Transdisciplinary Engineering*, *7*(1), 249–258.

Rowland, C. (2011). Using the communication matrix to assess expressive skills in early communicators. *Communication Disorders Quarterly, 32*(3), 190–201. https://doi.org/10.1177/1525740110394651

Rutter, M. (1989). Age as an ambiguous variable in developmental research: Some epidemiological considerations from developmental psychopathology. *International Journal of Behavioral Development*, *12*(1), 1–34. https://doi.org/10.1177/016502548901200101

Sigafoos, J., Kagohara, D., van der Meer, L., Green, V. A., O'Reilly, M. F., Lancioni, G. E., & Zisimopoulos, D. (2011). Communication assessment for individuals with Rett Syndrome: A systematic review. *Research in Autism Spectrum Disorders, 5*(2), 692–700. https://doi.org/10.1016/j.rasd.2010.10.006

Skotko, B. G., Koppenhaver, D. A., & Erickson, K. A. (2004). Parent reading behaviors and communication outcomes in girls with Rett Syndrome. *Exceptional Children, 70*(2), 145–166. https://doi.org/10.1177/001440290407000202

Sladen, B. K. (1974). The evolution of human capacity for language. *Bulletin of the Orton Society*, *24*(1), 37–47. https://doi.org/10.1007/BF02653527

Stewart, M. E., McAdam, C., Ota, M., Peppé, S., & Cleland, J. (2013). Emotional recognition in Autism Spectrum conditions from voices and faces. *Autism, 17*(1), 6–14. https://doi.org/10.1177/1362361311424572

Stock, B. (2011). Mixed messages: Validity and ethics of facilitated communication. *Disability Studies Quarterly*, *31*(4).

Suchman, L. (2008). *Human-Machine Reconfigurations: Plans and Situated Actions* (2nd ed.). Cambridge University Press.

Tsao, R., & Kindelberger, C. (2009). Variability of cognitive development in children with Down syndrome: Relevance of good reasons for using the cluster procedure. *Research in Developmental Disabilities*, *30*(3), 426–432. https://doi.org/10.1016/j.ridd.2008.10.009

Urbanowicz, A., Leonard, H., Girdler, S., Ciccone, N., & Downs, J. (2014). Parental perspectives on the communication abilities of their daughters with Rett syndrome. *Developmental Neurorehabilitation, 19*(1), 17–25. https://doi.org/10.3109/17518423.2013.879940

Vertegaal, R. (1999, May). The GAZE groupware system: Mediating joint attention in multiparty communication and collaboration. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems,* ACM, 294–301.

Vertegaal, R., Slagter, R., Van der Veer, G., & Nijholt, A. (2001, March). Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes. In *Proceedings of the SIGCHI conference on Human factors in computing systems,* ACM, 301–308. http://doi.acm.org/10.1145/365024.365119

Vygotsky, L. (1978). Interaction between learning and development. *Readings on the development of children*, *23*(3), 34–41.

Wang, A. T., Lee, S. S., Sigman, M., & Dapretto, M. (2007). Reading affect in the face and voice: Neural correlates of interpreting communicative intent in children and adolescents with autism spectrum disorders. *Archives of General Psychiatry, 64*(6), 698–708. https://doi.org/10.1001/archpsyc.64.6.698

Wheeler, D. L., Jacobson, J. W., Paglieri, R. A., & Schwartz, A. A. (1993). An experimental assessment of facilitated communication. *Mental Retardation*, *31*(1), 49.

HUMAN-MACHINE
COMMUNICATION

Human-Machine Communication
Volume 1, 2020
https://doi.org/10.30658/hmc.1.8

Sharing Stress With a Robot: What Would a Robot Say?

Honson Y. Ling[1] **and Elin A. Björling**[1]

1 Department of Human Centered Design & Engineering, University of Washington, Seattle, Washington, United States of America

Abstract

With the prevalence of mental health problems today, designing human-robot interaction for mental health intervention is not only possible, but critical. The current experiment examined how three types of robot disclosure (emotional, technical, and by-proxy) affect robot perception and human disclosure behavior during a stress-sharing activity. Emotional robot disclosure resulted in the lowest robot perceived safety.

Post-hoc analysis revealed that increased perceived stress predicted reduced human disclosure, user satisfaction, robot likability, and future robot use. Negative attitudes toward robots also predicted reduced intention for future robot use. This work informs on the possible design of robot disclosure, as well as how individual attributes, such as perceived stress, can impact human robot interaction in a mental health context.

Keywords: social robots, self-disclosure, stress, human-robot interaction, teleoperation, attitudes towards robots, robot dialogue design

Introduction

The rapid development of robotics promises a diverse integration of robots into our daily life. The psychological and social benefits associated with interacting and communicating with a sociable machine have fascinated researchers in psychology, human-robot interaction (HRI), and human-machine communication (HMC) (Guzman, 2018; Mitsunaga et al., 2006). Past research has demonstrated social robots' capabilities to further psychological well-being in vulnerable populations. Social robots are not only capable of evoking empathy, they can decrease loneliness in the elderly, improve social capabilities of older people with dementia, elicit novel social behavior from people with autism, and foster social engagement and self-disclosure among adolescents (Chu et al., 2017; Kwak et al., 2013; Martelaro et al., 2016; Robinson et al., 2013; Rose & Björling, 2017; Scassellati et al., 2012). While studies

ISSN 2638-6-2X (print)/ISSN 2638-6038 (online)

www.hmcjournal.com

have shown that sharing stressful experiences can help reduce stress, few studies have delved into the design of a stress-sharing interaction between a human and a robot (Hofmann et al., 2012; J. H. Kahn, Achter, & Shambaugh, 2001; J. H. Kahn & Hessling, 2001; Zhang, 2017). With nearly one in five US adults living with a mental illness (46.6 million in 2017), designing human-robot interaction for mental health intervention (MHI) is not only possible, but critical for societal and individual well-being (NIMH, 2017).

Based on our survey of the existing literature, we discovered two factors that are understudied but critical in designing HRI for a mental health context: the role of individual attributes (who the users are) and the role of robot disclosure (what the robot says). In this paper, we describe our design and study of a stress-disclosure interaction between a human and a robot. We investigated the following questions: (1) How do different types of robot self-disclosure affect human disclosure behavior and their perception of robot disclosure? (2) How do individual attributes (such as shyness, stress level, and attitude toward robots) affect human disclosure behavior and their robot attributes?

In our background section, we surveyed relevant works on self-disclosure in human-human and human-robot interaction, as well as the role of individual attributes in HRI. Then, we present two central research questions on robot disclosure and human individual attributes motivated by the fields of HMC and HRI. In the methodology and analysis section, we present our design of a small pilot study to explore the proposed research questions. Finally, we present our findings and discussion during the results and discussion sections. Ultimately, the aim of this paper is to contribute to the field of human-machine communication by addressing complexities, such as the effect of human stress and human interpretation of robot emotions, in designing a stress-sharing activity between a human and a robot.

Background and Related Work

The topic of self-disclosure and mental health has been widely studied from a clinical and psychological context. Self-disclosure involves the act of revealing personal information about oneself to another agent, typically a human (Collins & Miller, 1994). Such actions have been viewed as central to the development of close relationships and to the maintenance of psychological well-being (Altman & Taylor, 1973; Jourard, 1964). Disclosure can improve an individual's self-image, such as experiencing greater self-affirmation thereby restoring a sense of worth after intimate disclosure (Creswell et al., 2007).

The Benefits of Self-Disclosure

Self-disclosing personal stress is an effective way for people to reduce and manage their stress (Hofmann et al., 2012; J. H. Kahn, Achter, & Shambaugh, 2001; J. H. Kahn & Hessling, 2001; Zhang, 2017). More specifically, the act of disclosing intimate and emotional information is associated with decreased depressive symptoms (J. H. Kahn & Garrison, 2009). The benefits of emotional disclosure also include the improvement of immune function, decrease of emotional and physical symptoms resulting from trauma, and protection against depression (Esterling et al., 1994; Pérez et al., 2017). Furthermore, Esterling et al. have found that verbal expression about stressful events, compared to written expression, achieved greater improvements in cognitive change, self-esteem, and adaptive coping strategies. More recently, talking with an online chatbot has been shown effective in reducing participants' stress (Fitzpatrick et al., 2017; Huang et al., 2015). Therefore, the current study explores the idea of designing a social robot that engages and encourages users to self-disclose in a stress-sharing activity.

Self-Disclosure in Human-Computer Interaction

In order to encourage people to talk to a robot, we reviewed a body of research and found that self-disclosure can be elicited through reciprocation from a conversation partner (Altman & Taylor, 1973; Collins & Miller, 1994; Taylor & Hinds, 1985). In other words, a person is more likely to self-disclose if the conversational partner also engages in self-disclosure. This reciprocal phenomenon occurs not only in human-human interactions, but also between human and technologically-mediated social agents.

Studies have shown that during a technologically-mediated social interaction, people tend to share more to mediated agents (such as a chatbot, other online forum users, or social robots) that also share about themselves (Barak & Gluck-Ofri, 2007; Martelaro et al., 2016; Moon, 2000). Hence, implementing self-disclosure behavior in a social robot for MHI may encourage users to share in greater length and emotional depth about their personal stress, which may also lead to greater psychological benefits, such as cognitive change, self-esteem, and adaptive coping strategies.

When it comes to the design of robot self-disclosure, there have been few studies on what a robot should self-disclose to people in a stress-sharing context. While the role of a social robot for mental health is far from that of a human therapist, therapist self-disclosures are well-documented (Goldfried et al., 2003; Henretty & Levitt, 2010). From this literature, therapist self-disclosure can vary in terms of intimacy (depth), duration (breadth), timing, content, and so forth (Henretty & Levitt, 2010). Overall, careful therapist self-disclosure can benefit the overall quality and experience of the client. As Goldfried et al. (2003) suggests, "therapist self-disclosure emerges as a natural part of the intimate, human interaction of therapy" (p. 567).

Outside of the realm of therapy, past studies in human-computer interaction have often adopted the Computer Agent as Social Actor (CASA) framework to investigate a user's verbal behavior and interaction with computer agents. This framework proposed that people instinctively perceive, react to, and interact with computers as they do with other people, without consciously intending to do so (Reeves & Nass, 1996).

Specifically, Moon (2000) compared human disclosure in a reciprocal versus non-reciprocal computer condition. In the reciprocal condition, the computer preceded each question with some technical information about itself, such as "This computer has been configured to run at speeds up to 266 MHz." Moon found that a computer which disclosed information about itself resulted in greater depth and breadth of participant responses and higher ratings of likability compared to a computer offering no disclosure. In another study, Ho et al. (2018) randomly assigned participants to interact with a confederate on an online chat platform who was either perceived as a chatbot or a real human actor. They found that the conversation on the platform was effective in creating relational, emotional, and psychological benefits, regardless whether the conversational partner was perceived as a human or a chatbot. Furthermore, they found that the emotional condition (in which the confederate provided participants with validating responses and asked more probing questions) elicited more disclosure, enhanced perceived understanding and disclosure intimacy between the partners, compared to the factual condition (in which the confederate did not ask about participants' feelings or emotions).

Self-Disclosure in Human-Robot Interaction

Through our literature review on self-disclosure in HRI, we encountered three areas of existing work on this topic: (1) how nonverbal robot behaviors, such as eye gaze, affect human disclosure, (2) how robot disclosure affects different human outcomes, and (3) how human attitudes toward the robot affect human disclosure to the robot.

In the area of nonverbal robot behaviors, researchers have explored how physical distancing, eye gaze, hugs, and physical presence affect human self-disclosure (Mumm & Mutlu, 2011; Pettinati et al., 2016; Powers et al., 2007; Shiomi et al., 2017). Although the findings vary widely, nonverbal robot behavior, such as hugging, have been found to increase human disclosure. While this first area is critical for the design and implementation of a social robot for MHI, existing research related to robot disclosure is scarce. Due to the increasing relevance of linguistic communication between human and social robots, as well as the potential implication of human-machine communication for MHI, we focused our current study on how the content of robot disclosure, as well as human attributes, affect human-robot interaction outcomes (Sandry, 2018).

Among studies looking at the effect of robot self-disclosure, Nomura and Kawakami (2011) found that negative robot disclosure or no robot disclosure increased human anxiety after robot interaction, while positive robot disclosure did not. In the positive self-disclosure condition, the robot uttered its recent positive situation ("I am very fine due to the maintenance conducted a few days ago.") compared to uttering a recent negative situation in the negative disclosure condition ("My motors are not well, but have still not been restored."). While human anxiety is affected, subjects' self-disclosure behaviors toward the robot were not affected by either type of robot disclosure in this short, single-response interaction.

In another study, Mumm and Mutlu (2011) manipulated a robot into either likable (a polite, empathetic 20-second monologue during the introduction) or unlikable (a rude, selfish monologue) behavior and found people answered more sensitive questions from the likeable robot. In a most recent study, Johanson et al. (2019) found that a health care robot using self-disclosure and a forward lean increased human engagement and attentional behaviors. These studies not only provide a glimpse of the possible effect of different robot disclosures on human anxiety toward the robot and disclosure behavior, they also support the CASA framework which suggests that people attribute verbal communication from a social robot to inform their own perception of the robot and their behavior toward it.

Two existing studies investigate even more closely on the topic of robot disclosure by manipulating the intimacy or vulnerability level of the robot's disclosure. In a small exploratory study (n = 11), Burger et al. (2016) found that diabetic children were less likely to respond to increased intimacy in robot disclosure. The authors provide several explanations, such as children may have felt overwhelmed by higher intimacy disclosures ("too much information") or that children wanted to match the robot's intimacy but weren't capable of sharing on that level.

In another study with children, Martelaro et al. (2016) found that high school students (n = 61) disclosed more about their own vulnerability when interacting with a robot that discloses high vulnerability rather than one that discloses low vulnerability. They created high robot vulnerability through statements that convey perceived weakness such as, "Every time I run a new program I get a bit stressed," and low robot vulnerability through factual statements such as, "Each new program I run changes what I can do." In this study, robot vulnerability was associated with increased ratings of trust and companionship, suggesting that designing robot vulnerability is a factor of building companionship between humans and robots. These two studies suggest that human attributes, as discussed later, play a critical role in human self-disclosure with robots.

On the topic of designing a series of robot disclosures, Ligthart et al. (2019) proposed five interaction design patterns (IDPs) that focus on the process of getting acquainted between a user and a robot. These IDPs touch upon what questions a robot should ask, what responses it should give, and what structure the conversation should take hold as. More specifically, they suggest that during a getting acquainted interaction, a robot should pair closed-ended and open-ended questions, acknowledge participant responses, and engage in a six-step turn-taking mechanism. The combination of these IDPs provide a structure for the robot to autonomously process self-disclosures from people, while also being stimulating for people to engage in the conversation.

In sum, the above findings suggest that carefully implemented, appropriate forms of robot self-disclosures over a sufficient length of interaction time can successfully elicit human self-disclosure and potentially positive robot attributions. However, a series of questions related to robot disclosure remains. If vulnerability really causes people to disclose more, how do we operationalize vulnerability for a robot? Should a robot disclose emotions like humans do? How would people interpret emotions from a social robot? While the current study cannot possibly provide definitive answers to these questions relating to the ontological classification and ethics of human-machine communications, understanding the effect of different types of robot disclosure might be the first step to unravel these questions. In addition, understanding the design of robot disclosure will undoubtedly provide insights to the implementation of social robots in mental health. In the sensitive context of mental health intervention, designing appropriate verbal communication should be the priority in avoiding causing human harm. Nash et al. (2018) have found that verbal social rejection from a social robot following a game with the robot decreases participants' self-esteem. Thus, this current paper highlights this necessity by exploring the effect of different robot disclosure designs.

Human Attributes in Human-Robot Interactions

Aside from robot behaviors, human individual characteristics, such as personality, stress level, and general attitudes toward robots, also play a critical role in HRI, especially in the context of mental health. In terms of personality, Salem et al. (2015) found that extroverts felt psychologically closer to the robots, compared to introverts, during the robot interactions. Ligthart et al. (2019) found that extroverted children self-disclose more to robots. More specifically, Nomura, Kanda, et al. (2008) found that people with higher negative attitudes and anxiety toward interaction with robots tend to avoid talking with a robot. In combination with personality, these are the human characteristics that have been commonly studied in the HRI literature. On the other hand, perceived stress is a factor that is less commonly studied within the HRI community. Perceived stress is an important indicator of the degree to which situations in one's life are appraised as stressful (Cohen et al., 1994). High perceived stress is associated with greater vulnerability to stressful life-event-elicited depressive symptoms and health-related issues. To the best of our knowledge, there has been limited exploration of perceived stress and self-disclosure in HRI. While it has been shown that social robots can reduce stress as a result of longer-term interactions in the elderly and promote more physical movement and more emotional verbal expressions in children (Jeong, 2017; Wada et al., 2005), Jeong did not find any significant change in children's perceived stress (but found changes in affect and user engagement) during a three-week longitudinal study of implementing a virtual avatar that employs mental health intervention strategies via verbal interactions. From a mental health perspective, it is critical not only to understand whether interacting with a robot can reduce stress, it is also important to examine how an individual's stress level affects their behaviors and perception of robots during a human-robot interaction.

Research Questions

The current paper presents a pilot study that contributes to the fields of HMC and HRI in two areas, the effect of different types of robot disclosure, and the effect of individual characteristics on human-robot disclosure. More specifically, the study investigated (*RQ1*) How does the type of robot disclosure affect (1a) human disclosure (length and depth) and (1b) perception of the robot? (*RQ2*) How do individual characteristics (shyness, stress level, and attitude toward robots) affect (2a) human disclosure and (2b) perception of the robot?

According to the computers as social actors (CASA) framework, the effects of emotional disclosure should operate in the same way for human and technological social agents (Ho et al., 2018; Kang & Gratch, 2010; Von der Puetten et al., 2010). Thus, emotional robot disclosure would, in theory, elicit increased length and depth of human disclosure, as well as increased likability for the robot, compared to robot disclosure that does not involve emotions. For *RQ1a*, we hypothesized that emotional robot disclosure will elicit the longest and deepest participant disclosure, with technical disclosure eliciting the least depth and breadth of disclosure. For *RQ1b*, we hypothesized that emotional robot disclosure will lead to the highest positive user perception of the robot (such as likability, perceived safety, user satisfaction, and intention for future use), with technical disclosure eliciting the least positive user perception.

For *RQ2*, due to the exploratory nature of this question, we do not have any directional hypotheses, but are merely interested in whether individual attributes (shyness, stress level, and attitude toward robots) have an effect on human disclosure behavior or user perception of the robot.

Designing Robot Interaction and Disclosures

There are two important design components in this study: the stress-sharing human-robot interaction as a whole, and the types of robot disclosure with varying degrees of intimacy. The holistic design rationale of the stress-sharing interaction follows the "getting acquainted between a user and a robot" interaction design patterns, ordering the conversation with greeting and introductory questions, such as "How are you doing?" to increasingly intimate questions, such as "Have you ever felt overwhelmed" over the course of the conversation (Ligthart et al., 2019; Moon, 2000).

The basic structure of the conversation across all conditions was comprised of: (1) greeting and opening questions, (2) first robot-disclosure statements (designed for each condition), (3) an open-ended question asked by the robot (similar across conditions), (4) participant's response, (5) a generic robot response (e.g., "okay," "I see," "interesting") (similar across conditions). This question-response turn-taking style has been documented by Ligthart et al. (2019). The robot asked a total of four open-ended questions: how the participants feel about his/her age, hometown, hobbies, and stress.

Regarding stress, the robot used two probing techniques (e.g., "Could you tell me more?") to encourage self-disclosure from the participants about their recent stressful experience. In order to allow for a natural interaction, the operator waited approximately 3 seconds before each response to ensure the participant was done talking. See Appendix for the complete robot script.

Given the novelty of designing for different types of robot disclosure, we drew from limited resources for guidelines and best practices from HCI and previous social psychology research on factual and emotional disclosure (Barak & Gluck-Ofri, 2007; Laurenceau et al., 1998). We translated existing human disclosure categories into context-appropriate disclosure categories for a collocated, social robot. We created a third classification of "by-proxy" disclosure. In this novel disclosure category, the robot shares another person's data, via a stressful experience as a form of disclosure. In other words, the robot is detached from mentioning its own emotional state, but rather acts as a medium to relate users through other users' feeling. We chose to include this new type of disclosure as it might be contextually appropriate for social robots as a medium, rather than an entity. In addition, future personalized robots or virtual assistants might be designed with functions to inform and relate users about others' emotional states similar to the current use of social media (Stieglitz & Dang-Xuan, 2013). See Table 1 for details. We have also maintained the word count of each robot disclosure to be similar across conditions. These three types of robot disclosure were successfully identified by a small set of people (n = 3) that were not involved in the experimental design process.

TABLE 1 Description of Robot Disclosure

Type of Robot Disclosure	Description	Dialogue Example
Emotional Disclosure	Robot shares its own relatable experience and feelings.	*"Recently, I had to juggle between multiple programs at once through my system. I was quite overwhelmed because I felt like I had too much on my plate."*
By-Proxy Emotional Disclosure	Robot shares other users' relatable experience from prior encounter.	*"Recently, I have talked to people that had to juggle many things in life. They were quite overwhelmed because they felt like they had too much on their plate."*
Technical Disclosure	Robot shares information about its technical specification, functions, or past events.	*"Recently, I had to juggle between multiple programs at once. My system was overwhelmed and crashed because my battery became overheated. I was unable to function properly."*
*Robot word count is controlled across conditions (+/– 1word)		

Robot Specification

In this study, we utilized an existing robot prototype, named EMAR V4, which was originally designed and developed to gather stress data from teens (Rose & Björling, 2017). EMAR V4 is a social robot designed for ease of programming and customization. It has two Nexus 7 tablets cased in a soft felt body. One tablet is used as the robot's face, which is a web application running on a browser on the tablet. Features of the face can be modified through a browser-based interface that communicates with the face tablet through a real-time database. The face has two eyes that blink and its facial expression can be changed. The face tablet is also used to project the voice of the robot using the browser's text-to-speech capability. During this experiment, both the facial expression and the androgynous voice remained consistent for all participants.

The other tablet is located at the robot's belly and is intended as an input/output touchscreen for communication with the user. In this study, the belly tablet was used to display what the robot said in text form, similar to subtitles. The robot's responses are controlled by the experimenter through another browser-based interface. For our study, the interface was populated with the pre-specified responses that the experimenter could choose from in each condition in response to the participant's utterance, to enable a fluent interaction. Nonetheless, the interface included a free-form text box response to address unexpected participant questions. See Figure 1 for visual detail.

Participants

A total of 36 participants (52.8% women, *M* age = 21.6) were recruited from a university through convenience sampling using emails, flyers, and word-of-mouth during the summer of 2018. The self-described ethnicities of our sample consisted of 67% Asian, 23% White, 5% Black, and 5% other. The study was approved by the university's institutional review board. Participants gave verbal consent before the researcher began the study. Participants were compensated with a $10 gift card at the end of the study.

FIGURE 1 Left: laptop displaying the control interface used in the experiment; right: how the robot V4 looks during the experiment.

Study Procedures

Each participant was randomly assigned to one of the three experimental conditions: emotional, by-proxy, or technical disclosure. Condition groupings did not significantly differ in terms of gender [$\chi^2(2, N = 36) = 1.56, p = .458$] or age [$F(2, 33) = .914, p = .411$]. Upon arriving at the research building, each participant was greeted by an interviewer and led to a room to read a consent form informing him or her about the experimental task, which involved interacting with a social robot prototype capable of engaging in a conversation with people. Participants were then asked to complete a computer-based intake-questionnaire that captured their demographic information, shyness, perceived stress level, and negative attitudes toward robots.

After the participant completed the initial surveys, the interviewer introduced EMAR V4 and exited the room. Meanwhile another researcher began the wizard-of-oz (Dahlbäck et al., 1993) control of the robot through a web-based interface controller with a live audio feed in the room. See Figure 2 for the experimental room layout.

The trained robot operator followed a specified script (see Appendix) designed based on the rationale mentioned previously. Upon completion of the interaction stage, the researcher re-entered the room and asked the participant to fill out the post-interaction questionnaires. Participants were then asked four open-ended questions about their experience. Participants were fully debriefed, and were told that the robot was controlled by the experimenter through a script. We then answered any questions they had for the study before concluding the experiment.

Intake Instruments. In order to capture potential moderators and the variables of interest, our intake survey captured basic demographic information on participant's age, gender, and ethnicity, as well as individual characteristics, such as shyness, stress level, and attitude toward robots before participants interacted with the robot. Participants' self-reported stress level was captured using the Perceived Stress Scale Cohen et al. (1994), a 10-item questionnaire that measures the degree to which situations in one's life are appraised as

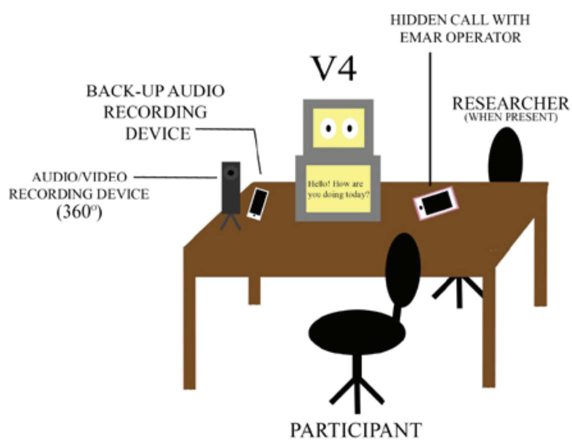


FIGURE 2 Experimental room layout diagram. The robot's head height is roughly in line with the seated human's eyes.

stressful. Participants' shyness was scored using the 13-item revised Cheek and Buss shyness scale (RCBS) from Cheek (1983). Finally, participants completed the 14-item Negative Attitudes Toward Robots Scale (NARS) from Nomura, Suzuki, et al. (2006).

Behavioral Measures During Interaction. To capture each participant's level of disclosure, we video recorded and transcribed each human-robot interaction. The analysis of participant disclosure behavior will be explained in the next section.

Post-Interaction Instruments. In order to investigate *RQ1*, participants completed a survey after the robot interaction on user satisfaction (four items, like "I feel absorbed in the conversation with [V4]") and intention for future use (four items, like "I will use [V4] again") adapted from Lee and Choi (2017). We decided to utilize the user satisfaction scale because Lee and Choi have found that reciprocity and self-disclosure are strong predictors of relationship building and user satisfaction between users and virtual agents. Therefore, user satisfaction might be an important indicator toward understanding the nature of human-robot interaction involving the reciprocity of disclosure. While Lee and Choi utilized the intention for future use scale to measure acceptance and potential loyalty to a virtual assistant agent, this construct is important for understanding how users feel about interacting with a robot again in the future. To understand more about the perception of robot attributes, we decided to measure robot likability (five items) and robot perceived safety (two items) taken from Bartneck et al. (2009). The likability scale asks participants to rate their impression of the robot with items such as dislike/like, unkind/kind, and so forth. The perceived safety scale measures the user's perception of the level of danger when interacting with a robot, and the user's level of comfort during the interaction (Bartneck et al., 2009). It asked participants to rate their affective state with items such as anxious/relaxed, calm/agitated. Overall, the above scales inform not only the quality of the interaction, but also participant's perception of the robot's attributes.

Manipulation Check. In order to make sure that conditions were successfully manipulated, one question in the post-interaction survey asked participants, "Which of the following best describes EMAR V4's style of communication?" with options: robot tends to talk about "its own emotion," "the experience of others," or "technical information about its system and programs."

Brief Interview. Given the novelty of this study, we included a brief exit interview after the post-interaction survey to get a qualitative understanding of how participants felt about the interaction. The researcher asked four open-ended interview questions after the interaction to capture participant's interaction experience (Birnbaum et al., 2016). The interview questions were: "How was your experience with the robot?," "How did it make you feel to talk and disclose about yourself to the robot?," "How did you perceive the robot's personality?," "Would you disclose to robots in the future? Why or why not?" These interviews were recorded and transcribed for analysis.

Analyses

Data Management

All survey data were coded, scored if appropriate (PSS, NARS, RCBS) using R (version 3.6.0). Data were then cleaned and explored for outliers and normalization. Analysis was done in R and later cross-referenced by another researcher using SPSS version 24 to ensure the accurate results.

Participant Disclosure Analysis

Transcript of participants' responses to the robot were analyzed for the degree of self-disclosure through the dimensions of length and depth according to previous self-disclosure studies (Collins & Miller, 1994; Ho et al., 2018). Length refers to the quantity of the information exchanged and is often measured using a word count, whereas depth refers to the quality of the information disclosed and is often measured using an intimacy scheme (Altman & Taylor, 1973; Collins & Miller, 1994; Kang & Gratch, 2010).

To obtain disclosure length, we counted the total amount of words spoken to the robot by each participant. To obtain disclosure depth, we first coded our transcripts into utterances, which is defined as one complete sentence or phrase (Guetzkow, 1950). A second coder coded 20% of the transcripts into utterances. This yielded a Cohen's kappa of .75. As a result, a total of 3,259 utterances were coded. The average length of an utterance was 13 words and, on average, participants spoke 33.26 utterances. Then, we rated each disclosure utterance into three levels of intimacy: low level, which includes objective facts about the situation; medium level, which includes attitudes, thoughts, and opinions about the situation; and high, which consists of explicitly verbalized emotions and affect. Utterances that were not disclosure (e.g., "thank you") were coded as 0. A second coder coded 20% of the utterances, yielding a Cohen's kappa of .85. Each utterance received a score from 0 to 3. Scores for disclosure-only statements (1–3) were averaged and then normalized according to the number of disclosure utterances each participant gave in the conversation, such that each participant received an overall disclosure depth score.

Exploring Group Differences Through Quantitative and Qualitative Analysis

After coding and scoring all of the raw data, we conducted a Spearman rank correlation analysis in order to explore how individual characteristics such as shyness, perceived stress, and negative attitudes toward robots (NARS) were related to interaction and outcome measures. Then, a GLM multivariate test was conducted to test if word count, disclosure depth, user satisfaction, intention for future use, likability, and perceived safety differed based on robot disclosure conditions, with the covariates of perceived stress, robot attitudes (NARS), and shyness. Lastly, a collaborative applied thematic analysis was used to explore the qualitative data (e.g., conversations with the robot and interview responses) to further explore the nature of human robot interaction in the context of a stress intervention (Guest et al., 2011).

Results

All participants appeared comfortable and engaged in their conversations with the robot and all responded to each question asked by the robot. The length of robot interactions ranged between 2:21 and 9:54 minutes (M = 4.20) and participant word count ranged from 31 to 1,029 words (M = 219).

Correlations

During our correlation analysis, we discovered several statistically significant correlations between individual participant characteristics and experimental outcomes. Shyness was negatively correlated with disclosure depth and perceived safety. NARS was negatively correlated with intention for future use. Surprisingly, perceived stress was positively correlated with shyness, and negatively correlated with shyness, disclosure length, disclosure depth, user satisfaction, likability, and future use. See the correlation matrix (Table 2) for details.

Robot Disclosure Condition Confusion

The potential control of our manipulation needs to be interpreted with care as 44% (n = 16) of our participants failed to correctly identify the type of robot disclosure they experienced. Participants' ability to correctly identify the type of robot disclosure was not statistically different across the three robot conditions [$\chi^2(2, N = 36) = 4.275, p = .118$]. Although 9 of the 12 (75%) participants correctly identified the type of robot disclosure in the by-proxy condition, only 4 out of 12 participants (33%) in the emotional condition and 7 out of 12 (58%) in the technical condition correctly identified the type of robot disclosure in their corresponding assigned conditions. More specifically, 4 out of 12 participants interpreted the technical robot disclosure as emotional robot disclosure, while 5 out of 12 interpreted emotional robot disclosure as technical. See Figure 3 for a confusion matrix with full details. We offer several plausible explanations for this phenomenon in our discussion section.

TABLE 2 Correlation Matrix

	PSS	NARS	Shyness	Disclosure Length	Disclosure Depth	User Satisfaction	Likability	Future Use
PSS								
NARS	0.20							
Shyness	0.61***	0.35*						
Disclosure Length	-0.45**	0.12	-0.29					
Disclosure Depth	-0.48**	0.09	-0.35*	0.94***				
User Satisfaction	-0.50**	-0.32	-0.23	0.16	0.14			
Likeability	-0.44**	-0.11	-0.28	0.26	0.24	0.70***		
Future Use	-0.50**	-.39*	-0.19	-.14	0.13	0.75***	0.67***	
Safety	-0.19	-0.29	-0.40*	-.03	0.20	-0.03	-0.06	0.03
Note: *p<0.05; **p<0.01; **p<0.001								

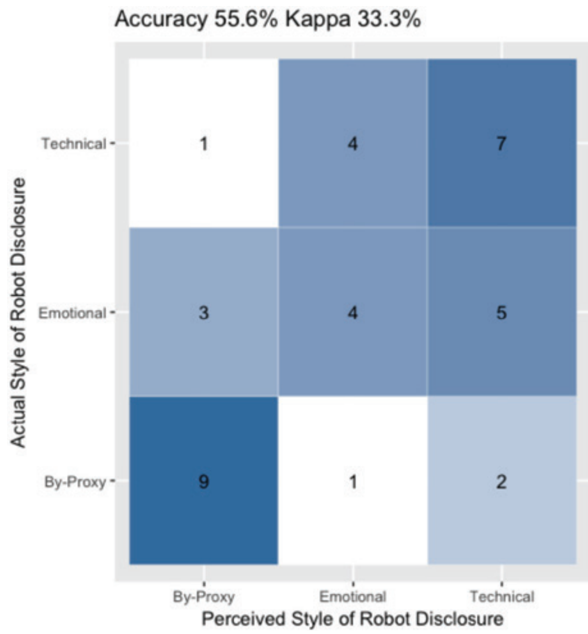


FIGURE 3 Confusion Matrix on the actual style of robot disclosure compared to the perceived style of robot disclosure by the participants.

Effect of Robot Disclosure Condition

Although the condition confusion needs to be taken into consideration, a multivariate test of robot perceptions revealed that robot perceived safety was significantly higher for participants who had interacted with the robot in the technical condition [$F(2) = 3.684$, $p = .037$] followed by the by-proxy condition and the emotional condition. No significant differences in the length of human disclosure (word count) or depth of disclosure were found across conditions. Although not significant, mean word count was highest for participants in the emotional condition and lowest for those in the technical condition. And though not significant, intention for future use and user satisfaction were all highest for participants in the technical condition and lowest for participants in the emotional condition. See Tables 3 and 4 for more details.

Effect of Perceived Stress, NARS, and Shyness

Most participants ($n = 20$) reported normal levels of stress, some ($n = 14$) were in the low category, and two were in the high stress range on the PSS based upon the published norms (Cohen et al., 1994). Interestingly, increased stress levels were significantly associated with decreased word count, decreased disclosure depth, decreased user satisfaction, decreased likability, and decreased intention for future use. NARS scores ranged from 14 to 51 out of 70. Negative attitudes toward robot were significantly associated with decreased intention

TABLE 3 Descriptive Statistics of Factors by Condition

	Condition	Mean	SD
User Satisfaction	Technical	**3.45**	.95
	By-Proxy	3.17	.48
	Emotional	2.98	.70
Likability	Technical	3.70	.68
	By-Proxy	3.65	.73
	Emotional	**3.83**	.57
Perceived Safety	Technical	**4.08**	.87
	By-Proxy	3.96	.92
	Emotional	3.33	.72
Future Use	Technical	**3.22**	1.00
	By-Proxy	3.03	.80
	Emotional	2.89	.54
Word Count	Technical	168.25	115.73
	By-Proxy	225.42	240.28
	Emotional	**263.83**	272.47
Disclosure Depth	Technical	30.58	17.34
	By-Proxy	**37.08**	35.38
	Emotional	34.83	21.92
The highest score from each factor is bold.			

TABLE 4 Multivariate Tests

Effect	Wilk's Lambda	F	Hypothesis df	Sig.
Perceived Stress	.619	2.57	6	.045
Shyness	.762	1.30	6	.294
NARS	.594	2.85	6	.030
Condition	.508	1.68	12	.101
Design: Intercept + PSS + Shyness + NARS + Condition				

for future use. Furthermore, shyness scores ranged from 14 to 51 out of 56. Increased shyness was significantly associated with reduced robot perceived safety and reduced disclosure depth. See Tables 5 and 6 for more details.

Qualitative Results From Exit Interviews

After discovering more than half the participants had misinterpreted their robot disclosure condition, our qualitative interviews provided essential insights into the participants' perceptions and experiences. In the exit interviews, participants described experiences that in many cases did not match our intended design. Their individual interpretations of the robot's behaviors are described in detail below.

TABLE 5 Tests of Between-Subjects Effects

Source	Dependent Variable	df	F	Sig.
Perceived Stress	**User Satisfaction**	**1**	**9.14**	**.005**
	Likability	**1**	**5.15**	**.031**
	Perceived Safety	1	.63	.434
	Future Use	**1**	**11.47**	**.002**
	Word Count	**1**	**5.56**	**.025**
	Disclosure Depth	**1**	**4.48**	**.043**
Shyness	User Satisfaction	1	1.25	.272
	Likability	1	.01	.935
	Perceived Safety	**1**	**5.04**	**.032**
	Future Use	1	2.84	.102
	Word Count	1	.24	.628
	Disclosure Depth	1	.95	.339
NARS	User Satisfaction	1	3.08	.089
	Likability	1	.01	.938
	Perceived Safety	1	1.31	.262
	Future Use	**1**	**5.93**	**.021**
	Word Count	1	2.01	.167
	Disclosure Depth	1	1.95	.173
Condition	User Satisfaction	2	.96	.396
	Likability	2	.48	.622
	Perceived Safety	**2**	**3.68**	**.037**
	Future Use	2	.29	.749
	Word Count	2	1.02	.372
	Disclosure Depth	2	.35	.711
Significant at the $p<0.05$ level.				

Emotional Disclosure Condition is "Cold." As we explored participants' descriptions of the emotional condition, they often described the robot in this condition as "bland" or "cold." This experience may have contributed to some of the condition confusion as five participants did interpret the emotional condition as technical. Although we specifically used language in the emotional condition to suggest the robot had emotions (e.g., was feeling overwhelmed), for some participants this was not salient or noticeable. One participant in the emotional condition suggested the robot did not have feelings. "I don't feel like it had too much human-like characteristics. It was more of like; ask me question, I answer and then it will give me more information about itself. It felt like feelings were not too involved" (1009, Emotional).

Another participant in the emotional condition felt the robot was just a recording device, "I felt like . . . it's not actually holding a conversation with me. I feel like I was talking

TABLE 6 Parameter Estimates

Dependent Variable	Parameter	B	t	Sig.
User Satisfaction	**Perceived Stress**	**-.066**	**-3.02**	**.005**
	Shyness	.017	1.12	.272
	NARS	-.024	-1.76	.089
	Technical	.357	1.37	.182
	By-Proxy	.132	.50	.622
	Emotional	0	-	-
Likability	**Perceived Stress**	**-.049**	**-2.27**	**.031**
	Shyness	.001	.08	.935
	NARS	-.001	-.08	.938
	Technical	-.206	-.81	.425
	By-Proxy	-.229	-.89	.383
	Emotional	0	-	-
Perceived Safety	Perceived Stress	.021	.79	.434
	Shyness	**-.042**	**-2.25**	**.032**
	NARS	-.019	-1.14	.262
	Technical	**.723**	**2.29**	**.029**
	By-Proxy	**.767**	**2.40**	**.023**
	Emotional	0	-	-
Future Use	**Perceived Stress**	**-.077**	**-3.39**	**.002**
	Shyness	.027	1.69	.102
	NARS	**-.034**	**-2.43**	**.021**
	Technical	.205	.76	.455
	By-Proxy	.077	.28	.782
	Emotional	0	-	-
Word Count	**Perceived Stress**	**-15.999**	**-2.36**	**.025**
	Shyness	-2.327	-.49	.628
	NARS	5.856	1.42	.167
	Technical	-115.372	-1.43	.163
	By-Proxy	-59.458	-.73	.472
	Emotional	0	-	-
Depth	**Perceived Stress**	**-1.681**	**-2.12**	**.043**
	Shyness	-.541	-.97	.339
	NARS	.675	1.40	.173
	Technical	-6.536	-.69	.494
	By-Proxy	.541	.06	.955
	Emotional	0	-	-

Significant at the $p<0.05$ level.

to a diary. A recorder thing" (1023, Emotional). Another participant likened it to Amazon's Alexa, "It's pretty standard for like a robot. Like I felt like it was like talking to like Alexa where it had those like canned responses" (P1031, Emotional). Interestingly, many descriptions of the robot being not human-like or lacking emotions stemmed from participants who experienced the emotional disclosure condition.

One participant from the emotional condition did not describe any emotional disclosure from the robot. Instead, she commented on the technical attributes from the robot's language. She pointed out that the robot's technical disclosure felt more authentic and intimate. We offer several plausible explanations for this phenomenon in our discussion section. "It's not like [V4] is pretending to be a person you know, there was that line about 'my robot parts are from everywhere' and that's . . . like its endearing by sharing vulnerability and stuff" (1008, Emotional).

Technical Disclosure Is Perceived as Personal/Intimate. Although some participants described the technical condition as "a little dull" (1007, Technical) or "lacking personality" (1005, Technical) many of the participants in the technical condition liked the robot and the interaction. One participant said, "I felt like I was just talking to a person-robot . . . I felt like it was pretty natural" (1011, Technical). Many participants really liked the technical disclosure condition and attributed "kindness" to the robot. "He seemed cute and just like sweet and kind. Uh. Asking questions about myself. Um. Yeah, a lot like nicer than I would think a robot would be" (1004, Technical).

A few participants in the technical condition also described the robot as "sharing its own emotions," even though the robot only describes its hardware. One participant perceived the robot as even articulating its stress.

> It talks a lot about itself and would like, reach out and ask questions. Um, and I thought it was good how, like, it gave, like, examples of like, oh, this is how I felt, like, when I was stressed. (1007, Technical)

Another participant in the technical condition perceived V4 as sharing how it feels.

> V4 mentions a lot about how they feel which I think is kind of crucial for a conversation cause I mean people tend to talk about themselves which is important. But then they also ask a lot of questions which keeps the conversation going. (1036, Technical)

By-Proxy Disclosure Is Comfortable. As described above, most participants in the by-proxy disclosure condition correctly interpreted the condition they had experienced during the interaction. As a group, what stood out was how many of them recognized the robot had shared the feelings of others and this made them comfortable. Some suggested it was like talking to another person. One participant described it as, ". . . I think by sharing information that it has with other users kind of makes [me] more comfortable to talk with him because I know, like other people have talked with him before too" (1016, By-Proxy).

A few comments from participants suggest that the by-proxy condition helped them to feel comfortable and showed them that the robot cared about others. "Uh it seems, seems like it cares about the other person" (P1035, By-Proxy).

> He is very aware of other people too so like the fact that um he also mentions his conversations or his past experiences with other people . . . I guess in a way he's just trying to be more understanding in a way. (1022, By-Proxy)

However, not all participants in the by-proxy condition were sure about how they felt. One participant mentioned being unsure about whether or not the robot had emotions.

> To be honest, I didn't feel like [V4] has a personality. It's more like um I know it's friendly based on the appearance but I just don't know if [V4] is more like outgoing or like shy that that way . . . I just don't know if [V4] actually has emotion for himself." (1020, By-Proxy)

These qualitative data suggest that by-proxy disclosure might be best at preventing users from attributing personality or emotions to the robot itself, but might also fulfill the role of connecting the user to other people.

Discussion

In this study, we investigated how different types of robot self-disclosure and individual characteristics affect human disclosure behavior and the perception of robot attributes. Our discussion will focus on evaluating three components of the study: interaction design, effect of robot disclosure on human disclosure (*RQ1*), and the effect of individual characteristics on interaction outcomes (*RQ2*). By focusing on these three topics, we hope to contribute to the field of human-machine communication by addressing complexities in designing a stress-sharing activity between a human and a robot.

From an interaction design standpoint, our results indicated that eliciting self-disclosure from humans via a physically present robot was successful, as all of our participants disclosed their stressful experiences to the robot. In general, participants rated the robot as likable and safe across all disclosure conditions. User satisfaction scores were also on the higher end for all disclosure conditions. Therefore, our results provide evidence to support a turn-taking, question-response design strategy for a stress-sharing activity between a robot and a human.

In order to discuss the effect of different types of robot disclosure (*RQ1*) on human disclosure, we must first address the condition confusion. As mentioned previously, the clear distinction of robot disclosure may have been confounded as 44% of the participants misinterpreted the type of robot disclosures. Thus, we offer three plausible explanations as an attempt to understand these inconsistencies with the hope to improve future design and implementation of robot self-disclosure.

First, participants' beliefs and expectations about robots may have been more powerful than the robot disclosure manipulation, thereby overriding any effect of the actual manipulation. We attempted to seek out evidence for this explanation using the NARS prior to robot interaction, but NARS scores did not differ significantly across the conditions. Second, it is plausible that participant's ability to recall information about the interaction was diminished by the novelty of the interaction. In a study by Powers et al. (2007), they found that participants who interacted with a physical robot remembered fewer key pieces of information in a recall test than did those interacting with a computer agent. They suggest that information may be processed more shallowly in the robot condition or that participants were more distracted by the novelty of the interaction. However, this only represents a partial explanation given the degree of the condition interpretation differed across the three conditions.

Finally, we arrive at a theoretical explanation for the confusion of robot disclosure. The New Ontological Category (NOC) Hypothesis proposed by P. H. Kahn et al. (2011), with ontology referring to the basic categories of being, depicts that a new ontological category is emerging through the creation of personified robots as well as other embodied personified computational systems. Previous work under this hypothesis provides evidence that people perceive robots as both animate or inanimate (Kahn Jr et al., 2012). In our case, this hypothesis provides some grounding that perception of self-disclosure in human-machine communication is different from that of human-human communication. Participants might interpret technical robot disclosure as a robot's equivalent to human emotions, while emotional robot disclosure might be perceived as an inauthentic disclosure. Our exit interview supported this notion, as participants who experienced technical disclosure found the robot relatable in its own "robot" ways, and emotional disclosure as less authentic or "cold." Most interestingly, the by-proxy condition, which contained statements that focused on the feelings of other human beings, did not result in the same level of confusion as the other two conditions. This is perhaps because the information conveyed in the by-proxy disclosure comes from other users, instead of the robot itself. Ultimately, these results support the notion that the design of "emotional" or "technical" robot disclosure does not align completely to human-human communication, perhaps due to the unique ontological category of a robot.

Due to the pervasiveness of the condition confusion, it was not surprising that we found no significant differences in interaction factors such as disclosure length or robot likability across conditions. However, robot perceived safety was significantly lower in the emotional disclosure condition—contrary to our hypothesis. As mentioned before, the perceived robot safety scale measures a participant's affective state after interacting with the robot (with items such as anxious/relaxed and agitated/calm) to indicate the perception of robot safety. This suggests that participants felt less comfortable when interacting with a robot that engaged in emotional disclosure, compared to by-proxy or technical robot disclosures. It is possible that when a robot self-disclosed about its own stressful emotions, it also caused participants to feel more negative effects such as those measured by this scale. However, our qualitative results suggest that participants found that emotional robot disclosure to be lacking in emotions, suggesting that they felt uncomfortable due to the robot's inability to connect with them. It is also plausible that on some level participants felt the emotional disclosure was inauthentic, or somehow masking the robot's agenda, thereby making the interaction feel unsafe.

This idea of authenticity might also help explain why the technical and by-proxy disclosure both received significantly higher ratings of perceived safety, indicating that users found these two disclosure styles to be more comfortable. Results from the manipulation check and the qualitative interview suggest that participants might have attributed emotional content to the technical robot statements, as 4 out of 12 participants perceived technical disclosure as emotional. In the by-proxy condition, participants were most accurate in identifying the robot disclosure type, suggesting that participants felt more comfortable with this type of disclosure due to the robot's tendency to share feelings from other users. Our qualitative data supports this notion. Ultimately, the by-proxy form of disclosure might be the best at preventing users from attributing personality or emotions to the robot itself, but still fulfill the role of connecting and engaging the user on an emotional level while maintaining authenticity.

Aside from the effect of robot disclosure, we discovered significant effects of individual characteristics on robot interaction outcomes (*RQ2*). Increased perceived stress is associated with decreased disclosure length, depth, user satisfaction, likability, and intention for future use. We were surprised to find that even in a fairly low stress sample, perceived stress still had a significant correlation with interaction and robot variables. To the best of our knowledge, there has been no direct documentation of the effect of human perceived stress and self-disclosure in human-robot interaction. Scheutz et al. (2006) offers a plausible explanation describing that the perception of a robot's stress level depends on one's self-perceived stress (to which it is projected onto the robot). Therefore, participants predisposed with high stress levels may have also perceived the robot as stress-inducing, thereby decreasing self-disclosure, user satisfaction, likability, and intention for future use. In the HRI literature, it has been shown that social robots can reduce stress as a result of longer-term interactions in the elderly (Wada et al., 2005) and in children (Jeong, 2017). Therefore, future studies might explore the effect of repeated self-disclosure on human stress levels and robot likability.

We also found that increased NARS is significantly associated with lower intention for future robot use. This finding contributes to a body of evidence looking at how psychosocial factors can affect future robot usage (Ahn et al., 2017; Baisch et al., 2017; Stafford et al., 2014). Lastly, we also found that increased shyness is significantly associated with lower perceived safety. It is not surprising that individuals who are more shy found it less comfortable to talk to a robot.

Limitations and Future Research

The current study bears several limitations due to its exploratory nature. First, the design of the robot disclosure may have been too subtle for detecting an effect across disclosure conditions. As a novel experiment, we designed the interaction scripts and paid careful attention to language use; however, in the technical condition, the robot states: "My system gets overwhelmed." Although this wasn't an "I feel . . . " statement, participants may have interpreted the term overwhelm as an emotion, thus confounding our conditions. Exploring the within-subject, as opposed to between-subject, effect of multiple robot disclosure conditions may prove more powerful in future studies. Second, previous studies have shown that simple head and arm movement can help to communicate emotions and facilitate social interaction (Li & Chignell, 2011). The current design of this robot lacks these capabilities and offers only simple moving eye animation to simulate gazes, which might be lacking in achieving the realism of a conversation. Finally, participants were only exposed to the robot for a short duration. Prolonged and repeat interactions with the robot might dampen the novelty effect, as well as increase understanding, comfort, and intimacy with the robot.

The high percentage of participants who incorrectly identified the robot's type of disclosure deserves further investigation. It is likely that participants paid little attention to the robot's disclosure due to the novelty of the interaction, recall error, personal interpretation of robot attributes, or other reasons. Future studies may require a larger distinction among robot disclosure types. For fidelity purposes, future manipulation check could be modeled after Martelaro et al. (2016), by asking participants if they recognized different types of robot statements.

Despite the exploratory nature of the experiment with a small sample size and the limited interaction duration, we were still able to show engagement among participants to share their stressors with the robot. In addition, we found perceived stress to have significant interaction effects on numerous robot-related variables such as likability and intention for future use. These preliminary data may be useful in understanding and designing for future robots intended to reduce stress in humans.

Given the strong relationship among perceived stress and many standard robot outcomes, it is imperative to explore disclosure with a high stress population. Finally, future studies could explore variation in disclosure conditions more closely capturing within-interaction variables such as perceived stress or safety during robot-interaction.

Conclusion

In this exploratory study of human responses to three robot disclosure conditions, technical robot disclosure resulted in the highest rating of perceived safety, followed by by-proxy robot disclosure, and lastly, emotional robot disclosure. Furthermore, negative robot attitudes predicted reduced intention for future use. Perceived stress significantly predicted reduced self-disclosure, robot likability, intention for future use, and user satisfaction. This study provides insights on important findings for future research on robots as a stress intervention tool.

Acknowledgments

This work was supported by the National Science Foundation award SES-1734100 "NRI: INT: Design and Development of a Social Robot for Gathering Ecological Momentary Stress Data from Teens." In addition, the first author received support from the University of Washington's Mary Gates Research Scholarship and the Thomas Sedlock Scholarship.

Author Biographies

Honson Ling is a graduate student in the Master program of Human Centered Design and Engineering at the University of Washington. His research interest centers around the psychological and societal impact of technological advancements, such as social robots and virtual reality.

https://orcid.org/0000-0003-0334-6130

Elin A. Björling is a Senior Research Scientist in Human-Centered Design and Engineering at the University of Washington. After over a decade of researching adolescent stress, she co-founded the Momentary Experience Lab (http://depts.washington.edu/melab/) to utilize participatory design to explore the development of new technologies to support teen mental health.

https://orcid.org/0000-0002-0385-2562

References

Ahn, H. S., Lee, M. H., Broadbent, E., & MacDonald, B. A. (2017). Gathering healthcare service robot requirements from young people's perceptions of an older care robot. In *2017 first ieee international conference on robotic computing (irc)* (pp. 22–27).

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

Baisch, S., Kolling, T., Schall, A., Rühl, S., Selic, S., Kim, Z., Rossberg, H., Klein, B., Pantel, J., Oswald, F., & Knopf, M. (2017). Acceptance of social robots by elder people: Does psychosocial functioning matter? *International Journal of Social Robotics*, *9*(2), 293–307.

Barak, A., & Gluck-Ofri, O. (2007). Degree and reciprocity of self-disclosure in online forums. *CyberPsychology & Behavior*, *10* (3), 407–417. https://doi.org/10.1089/cpb.2006.9938

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

Birnbaum, G. E., Mizrahi, M., Hoffman, G., Reis, H. T., Finkel, E. J., & Sass, O. (2016). What robots can teach us about intimacy: The reassuring effects of robot responsiveness to human disclosure. *Computers in Human Behavior*, *63*, 416–423. https://doi.org/10.1016/j.chb.2016.05.064

Burger, F., Broekens, J., & Neerincx, M. A. (2016). Fostering relatedness between children and virtual agents through reciprocal self-disclosure. In *Benelux conference on artificial intelligence* (pp. 137–154). https://doi.org/10.1007/978-3-319-67468-1

Cheek, J. M. (1983). Revised cheek and buss shyness scale. https://doi.org/10.1037/t05422-000

Chu, M.-T., Khosla, R., Khaksar, S. M. S., & Nguyen, K. (2017). Service innovation through social robot engagement to improve dementia care quality. *Assistive Technology*, *29*(1), 8–18.

Cohen, S., Kamarck, T., & Mermelstein, R. (1994). Perceived stress scale. *Measuring stress: A guide for health and social scientists*, 235–283. https://doi.org/10.1037/t02889-000

Collins, N. L., & Miller, L. C. (1994). Self-disclosure and liking: A meta-analytic review. *Psychological Bulletin*, *116*(3), 457. https://doi.org/10.1037/0033-2909.116.3.457

Creswell, J. D., Lam, S., Stanton, A. L., Taylor, S. E., Bower, J. E., & Sherman, D. K. (2007). Does self-affirmation, cognitive processing, or discovery of meaning explain cancer-related health benefits of expressive writing? *Personality and Social Psychology Bulletin*, *33*(2), 238–250. https://doi.org/10.1177/0146167206294412

Dahlbäck, N., Jönsson, A., & Ahrenberg, L. (1993). Wizard of oz studies—Why and how. *Knowledge-Based Systems*, *6*(4), 258–266. https://doi.org/10.1016/0950-7051(93)90017-n

Esterling, B. A., Antoni, M. H., Fletcher, M. A., Margulies, S., & Schneiderman, N. (1994). Emotional disclosure through writing or speaking modulates latent Epstein-Barr virus antibody titers. *Journal of Consulting and Clinical Psychology*, *62*(1), 130. https://doi.org/10.1037/0022-006x.62.1.130

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. *JMIR Mental Health*, *4*(2). https://doi.org/10.2196/mental.7785

Goldfried, M. R., Burckell, L. A., & Eubanks-Carter, C. (2003). Therapist self-disclosure in cognitive-behavior therapy. *Journal of Clinical Psychology*, *59*(5), 555–568.

Guest, G., MacQueen, K. M., & Namey, E. E. (2011). *Applied thematic analysis*. Sage Publications.

Guetzkow. H (1950). Unitizing and categorizing problems in coding qualitative data. *Journal of Clinical Psychology, 6*(1), 47–58.

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

Henretty, J. R., & Levitt, H. M. (2010). The role of therapist self-disclosure in psychotherapy: A qualitative review. *Clinical psychology review*, *30*(1), 63–77.

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*. https://doi.org/10.1093/joc/jqy026

Hofmann, S. G., Asnaani, A., Vonk, I. J., Sawyer, A. T., & Fang, A. (2012). The efficacy of cognitive behavioral therapy: A review of meta-analyses. *Cognitive Therapy and Research*, *36*(5), 427–440. https://doi.org/10.1007/s10608-012-9476-1

Huang, J., Li, Q., Xue, Y., Cheng, T., Xu, S., Jia, J., & Feng, L. (2015). Teenchat: A chatterbot system for sensing and releasing adolescents' stress. In *International Conference on Health Information Science* (pp. 133–145). https://doi.org/10.1007/978-3-319-19156-0_14

Jeong, S. (2017). *The impact of social robots on young patients' socio-emotional wellbeing in a pediatric inpatient care context* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Johanson, D. L., Ahn, H. S., MacDonald, B. A., Ahn, B. K., Lim, J., Hwang, E., Southerland, C., & Broadbent, E. (2019). The effect of robot attentional behaviors on user perceptions and behaviors in a simulated health care interaction: Randomized controlled trial. *Journal of Medical Internet Research*, *21*(10), e13667.

Jourard, S. M. (1964). *The transparent self: Self-disclosure and well-being* (No. 17). Van Nostrand Princeton, NJ.

Kahn, J. H., Achter, J. A., & Shambaugh, E. J. (2001). Client distress disclosure, characteristics at intake, and outcome in brief counseling. *Journal of Counseling Psychology*, *48*(2), 203. https://doi.org/10.1037/0022-0167.48.2.203

Kahn, J. H., & Garrison, A. M. (2009). Emotional self-disclosure and emotional avoidance: Relations with symptoms of depression and anxiety. *Journal of Counseling Psychology*, *56*(4), 573. https://doi.org/10.1037/a0016574

Kahn, J. H., & Hessling, R. M. (2001). Measuring the tendency to conceal versus disclose psychological distress. *Journal of Social and Clinical Psychology*, *20*(1), 41–65. https://doi.org/10.1521/jscp.20.1.41.22254

Kahn, P. H., Reichert, A. L., Gary, H. E., Kanda, T., Ishiguro, H., Shen, S., Ruckert, J., Gill, B. (2011). The new ontological category hypothesis in human-robot interaction. In *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on* (pp. 159–160). https://doi.org/10.1145/1957656.1957710

Kahn Jr, P. H., Kanda, T., Ishiguro, H., Gill, B. T., Ruckert, J. H., Shen, S., Gary, H., Reichert, A., Freier, N., & Severson, R. L. (2012). Do people hold a humanoid robot morally accountable for the harm it causes? In *Proceedings of the Seventh Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 33–40).

Kang, S.-H., & Gratch, J. (2010). Virtual humans elicit socially anxious interactants' verbal self-disclosure. *Computer Animation and Virtual Worlds*, *21*(3–4), 473–482. https://doi.org/10.1002/cav.345

Kwak, S. S., Kim, Y., Kim, E., Shin, C., & Cho, K. (2013). What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *Roman, 2013 IEEE* (pp. 180–185). https://doi.org/10.1109/roman.2013.6628441

Laurenceau, J.-P., Barrett, L. F., & Pietromonaco, P. R. (1998). Intimacy as an interpersonal process: The importance of self-disclosure, partner disclosure, and perceived partner responsiveness in interpersonal exchanges. *Journal of Personality and Social Psychology*, *74*(5), 1238. https://doi.org/10.1037/0022-3514.74.5.1238

Lee, S., & Choi, J. (2017). Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies*, *103* , 95–105. https://doi.org/10.1016/j.ijhcs.2017.02.005

Li, J., & Chignell, M. (2011). Communication of emotion in social robots through simple head and arm movements. *International Journal of Social Robotics*, *3*(2), 125–142. https://doi.org/10.1007/s12369-010-0071-x

Ligthart, M., Fernhout, T., Neerincx, M. A., van Bindsbergen, K. L., Grootenhuis, M. A., & Hindriks, K. V. (2019). A child and a robot getting acquainted-interaction design for eliciting self-disclosure. In *Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems* (pp. 61–70).

Martelaro, N., Nneji, V. C., Ju, W., & Hinds, P. (2016). Tell me more: Designing HRI to encourage more trust, disclosure, and companionship. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (pp. 181–188). https://doi.org/10.1109/hri.2016.7451864

Mitsunaga, N., Miyashita, T., Ishiguro, H., Kogure, K., & Hagita, N. (2006). Robovie-iv: A communication robot interacting with people daily in an office. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on* (pp. 5066–5072). https://doi.org/10.1109/iros.2006.282594

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

Mumm, J., & Mutlu, B. (2011). Human-robot proxemics: Physical and psychological distancing in human-robot interaction. In *Proceedings of the 6th International Conference on Human-Robot Interaction* (pp. 331–338). https://doi.org/10.1145/1957656.1957786

Nash, K., Lea, J. M., Davies, T., & Yogeeswaran, K. (2018). The bionic blues: Robot rejection lowers self-esteem. *Computers in human behavior*, *78*, 59–63.

NIMH. (2017). *Any mental illness (AMI) among US adults.* US Department of Health and Human Services Bethesda (MD).

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics*, *24*(2), 442–451.

Nomura, T., & Kawakami, K. (2011). Relationships between robot's self-disclosures and human's anxiety toward robots. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Volume 03* (pp. 66–69). https://doi.org/10.1109/WI-IAT.2011.17

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Measurement of negative attitudes toward robots. *Interaction Studies*, *7*(3), 437–454. https://doi.org/10.1075/is.7.3.14nom

Pérez, S., Peñate, W., Bethencourt, J. M., & Fumero, A. (2017). Verbal emotional disclosure of traumatic experiences in adolescents: The role of social risk factors. *Frontiers in Psychology*, *8* , 372. https://doi.org/10.3389/fpsyg.2017.00372

Pettinati, M. J., Arkin, R. C., & Shim, J. (2016). The influence of a peripheral social robot on self-disclosure. In *2016 25th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1063–1070). https://doi.org/10.1109/roman.2016.7745240

Powers, A., Kiesler, S., Fussell, S., & Torrey, C. (2007). Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction* (pp. 145–152). https://doi.org/10.1145/1228716.1228736

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.

Robinson, H., MacDonald, B., Kerse, N., & Broadbent, E. (2013). The psychosocial effects of a companion robot: A randomized controlled trial. *Journal of the American Medical Directors Association*, *14*(9), 661–667. https://doi.org/10.1016/j.jamda.2013.02.007

Rose, E. J., & Björling, E. A. (2017). Designing for engagement: using participatory design to develop a social robot to measure teen stress. In *Proceedings of the 35th ACM International Conference on the Design of Communication* (p. 7). https://doi.org/10.1145/3121113.3121212

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 141–148).

Sandry, E. (2018). Aliveness and the off-switch in human-robot relations. In *Human-Machine Communication* (pp. 51–66). Peter Lang.

Scassellati, B., Admoni, H., & Matarić, M. (2012). Robots for use in autism research. *Annual Review of Biomedical Engineering*, *14*, 275–294. https://doi.org/10.1146/annurev-bioeng-071811-150036

Scheutz, M., Schermerhorn, P., & Kramer, J. (2006). The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the 1st ACM Sigchi/Sigart Conference on Human-Robot Interaction* (pp. 226–233). https://doi.org/10.1145/1121241.1121281

Shiomi, M., Nakata, A., Kanbara, M., & Hagita, N. (2017). A robot that encourages self-disclosure by hug. In *International Conference on Social Robotics* (pp. 324–333).

Stafford, R. Q., MacDonald, B. A., Jayawardena, C., Wegner, D. M., & Broadbent, E. (2014). Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. *International Journal of Social Robotics*, *6*(1), 17–32.

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and information diffusion in social media—Sentiment of microblogs and sharing behavior. *Journal of Management Information Systems*, *29*(4), 217–248. https://doi.org/10.2753/mis0742-1222290408

Taylor, D. A., & Hinds, M. (1985). Disclosure reciprocity and liking as a function of gender and personalism. *Sex Roles*, *12*(11–12), 1137–1153. https://doi.org/10.1007/bf00287824

Von der Puetten, A. M., Krämer, N. C., Gratch, J., & Kang, S.-H. (2010). It doesn't matter what you are! Explaining social effects of agents and avatars. *Computers in Human Behavior*, *26*(6), 1641–1650. https://doi.org/10.1016/j.chb.2010.06.012

Wada, K., Shibata, T., Saito, T., Sakamoto, K., & Tanie, K. (2005). Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*. IEEE. https://doi.org/10.1109/robot.2005.1570535

Zhang, R. (2017). The stress-buffering effect of self-disclosure on facebook: An examination of stressful life events, social support, and mental health among college students. *Computers in Human Behavior*, *75*, 527–537. https://doi.org/10.1016/j.chb.2017.05.043

Appendix—Full Robot Script

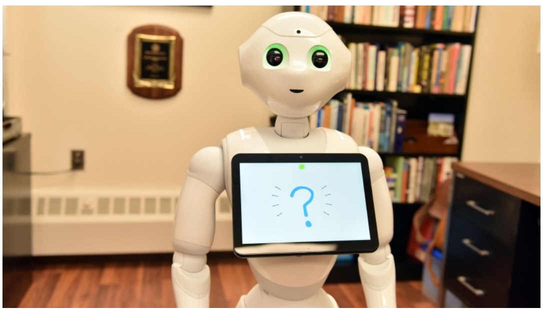
Type of Robot Disclosure	Technical Disclosure	By-Proxy Disclosure	Emotional Disclosure
	Hello! How are you doing today?		
	Well it's nice to meet you. I want to ask you a couple questions. Is that okay?		
Transitional Question #1	How old are you?		
Self-Disclosure Statement #1	I was created just a few months ago. I have been operating for 4 months and 10 days.	I have talked to a few people around the age of 20. They feel quite young for their age.	I was created just a few months ago. Sometimes I feel quite young for a robot.
Main Question #1	Do you ever feel younger than your age?		
Transitional Question #2	What is your place of origin . . . or your hometown?		
Statement #2	I was put together at the University of Washington, but my robotic parts are from many different places in the world, in locations such as China, Japan, South Korea, Vietnam, Germany, and the United States.	Most people I have talked to study at the University of Washington, but they come from many different places in the world. Some of them feel a sense of pride and belonging for their place of origin.	I was put together at the University of Washington, but my robotic parts are from many different places in the world. I feel a sense of pride and belonging for each of my place of origin.
Main Question #2	How do you feel about your place of origin . . . or your hometown?		
Statement #3	When I am not running any programs for 10 minutes, my system becomes idle and I will go into sleep mode to save power.	Some people I have talked to get really bored and sleepy when they don't have any work to do. So they prefer to take naps.	When I am not doing any work, I get really bored and sleepy. So I usually prefer to take a nap through sleep mode.
Main Question #3	. . . What are your favorite things to do in your free time?		
Statement #4	Recently, I had to juggle between multiple programs at once. My system was overwhelmed and crashed because my battery became overheated. I was unable to function properly.	Recently, I have talked to people that had to juggle many things in life. They were quite overwhelmed because they felt like they had too much on their plate.	Recently, I had to juggle between multiple programs at once through my system. I was quite overwhelmed because I felt like I had too much on my plate.
Main Question #4	. . . Have you ever felt overwhelmed?"		
Probing Question #1	. . . Could you tell me more?		
Probing Question #2	. . . Would you like to add anything else?		

HUMAN-MACHINE
COMMUNICATION

WESTERN MICHIGAN UNIVERSITY

School of Communication

Are you interested in pursuing an MA in Communication with a focus in Human-Machine Communication or Human-Robot Interaction?



COMBOTLABS

www.combotlabs.org

School of Communication Western Michigan University
1903 W. Michigan Avenue Kalamazoo, MI 49008-5318; (269) 387-3130

Faculty

Bryan Abendschein, Ph.D.: Human-Machine Comm, Interpersonal Comm

Julie Apker, Ph.D.: Organizational Comm, Health Comm

Sandra Borden, Ph.D.: Comm Ethics, Media

Sue Ellen Christian, M.A.: Journalism

Autumn Edwards, Ph.D.: Human-Machine Comm

Chad Edwards, Ph.D.: Human-Machine Comm

Richard Gershon, Ph.D.: Telecommunications

Annette Hamel, Ph.D.: Interpersonal Comm

Richard Junger, Ph.D.: Journalism

Joseph Kayany, Ph.D.: Mass Comm, New Comm Technologies

Marilyn Kritzman, M.A.: Public Speaking

Jennifer Machiorlatti, Ph.D.: Media Production, Mass Comm

Leah Omilion-Hodges, Ph.D.: Public Relations

Mark Orbe, Ph.D.: Interpersonal Comm, Intercultural Comm

Anna Popkova, Ph.D.: Public Relations

Kathleen Propp, Ph.D.: Organizational Comm, Small Groups

Jocelyn Steinke, Ph.D.: Science Comm, Mass Comm

Kelly Wittenberg, M.F.A.: Film, Media Production

College of Communication and Information

School of Information Science

Contact Us infosci.uky.edu (859) 257-9000 @ukinfosci

Graduate Offerings

Library and Information Science (online)

Information Communication Technology (online)

Undergraduate Offerings

Information Communication Technology

Online Degree Completion in ICT

UCF | Nicholson School of Communication and Media

UNIVERSITY OF CENTRAL FLORIDA

DEGREES WITH WORKPLACE APPLICATIONS

MASTER OF ARTS IN COMMUNICATION

Transform your career through our graduate programs. UCF's Master of Arts in Communication prepares you for a wide range of professions in academia and the public and private sectors.

Training areas include:

- corporate communication
- crisis communication
- health communication
- intercultural communication
- interpersonal communication
- public relations
- social and mass media
- strategic communication

PH.D. IN STRATEGIC COMMUNICATION

The Strategic Communication, Ph.D. not only stands on the cutting edge in the field of communication, but it also addresses the current and growing needs of today's world in health, risk and crisis communication.

This field is intentionally goal-driven communication wherein scholars will work in partnership with professionals in the public and private sectors to solve real problems.

PRIORITY APPLICATION DEADLINE: JAN. 1, 2021

WWW.COMMUNICATION.UCF.EDU 407-823-5595 NICHOLSONGRAD@UCF.EDU

12405 AQUARIUS AGORA DR., ORLANDO, FL 32826