Human-Machine Communication

Human-Machine Communication
Volume 4, 2022

Volume Editor

Leopoldina Fortunati, University of Udine (Italy)

Editor-in-Chief

Autumn Edwards, Western Michigan University (U.S.A.)

Associate Editors

Patric R. Spence, University of Central Florida (U.S.A.)
Chad Edwards, Western Michigan University (U.S.A.)

Editorial Board

CONTENTS

Framing the Psycho-Social and Cultural Aspects of 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

Abstract

In this introduction to the fourth volume of the journal *Human-Machine Communication*, we present and discuss the nine articles selected for inclusion. In this essay, we aim to frame some crucial psychological, sociological, and cultural aspects of this field of research. In particular, we situate the current scholarship from a historical perspective by (a) discussing humanity's long walk with hybridity and otherness, at both the cultural and individual development levels, (b) considering how the organization of capital, labor, and gender relations serve as fundamental context for understanding HMC in the present day, and (c) contextualizing the development of the HMC field in light of seismic, contemporary shifts in society and the social sciences. We call on the community of researchers, students, and practitioners to ask the big questions, to ground research and theory in the past as well as the real and unfolding lifeworld of human-machine communication (including what HMC *may* become), and to claim a seat at the table during the earliest phases in design, testing, implementation, law and policy, and ethics to *intervene* for social good.

Introduction

The community of human-machine communication (HMC) is showing great intellectual vitality. It attracts an increasing number of scholars drawn to the human-machine relationship as one of the central themes for understanding contemporaneity. The great effort that is required of us is to resist the simple fascination that machines exert on humans and assume the standpoint of critical witness to the sociocultural changes taking place under our gaze and advocate for social good. We introduce nine compelling and original articles which offer a substantial contribution specifically to Volume 4 (2022) of *Human-Machine*

Communication but also more generally to the field of HMC and neighboring disciplines. As a collection, these papers offer different lenses, ranging from narrativity (Prahl et al.), to pure philosophical inquiry (Gunkel), from content analysis (Richards et al.) to theory building (Gambino & Liu), from thorough reviews on specific topics such as robot sex (Dehnert) and supportive technologies (Beattie & High) to empirical explorations of robot penetration in three strategic sectors: (1) the domestic sphere including the trustworthiness of voice-based virtual assistants (Weidmüller), (2) the social presence of machine teachers in education (Kim et al.), and (3) the incorporation of exoskeletons in the factory (Kirkwood et al.). As editors, we are very grateful to these scholars for the variety of discourses they have introduced in their papers because this diversity in approach signals the presence of an enduring richness and difference of expression that is of vital importance to maintain in a scientific field. We are also very grateful to these contributors for the originality with which they have investigated themes, issues, and sectors of utmost relevance for the area of HMC.

Taking the Anticipatory Stance in Human-Machine Communication

This volume opens with an exemplary HMC case study conducted by Andrew Prahl, Rio Kin Ho Leung, and Alicia Ning Shan Chua. In "Fight for Flight: The Narratives of Human Versus Machine Following Two Aviation Tragedies," Prahl takes up the 737 MAX accidents to understand how the relationship between human and machine was constructed by professional aviators in an online discussion forum within the criticalities of the recent developments of automation in the aviation sector. This fascinating and insightful narrative analysis of professional pilots' discourse surrounding the tragedies touches on the constellation of technologizing factors that converge to constitute HMC in organizations (Evans, 2017; Fortunati & Edwards, 2021). As Prahl et al. maintain, by involving AI, sensors, judgment, and control, "the modern flightdeck is a quintessential example of the 'cluster' of technologies and processes that define modern workplaces where human-machine communication takes place" (Prahl et al., p. 29). Three encompassing narratives were identified in aviation professionals' discussion posts in a global online forum: *human versus machine as a zero-sum game* (a win for automation is a loss for human workers), *surrender to machines* (a belief in the inevitable triumph of automation processes), and *an epidemic of distrust* (pervasive loss of whole industry credibility). Prahl et al. raise important questions about the study of HMC involving machines that are not ostensibly designed to communicate, discuss implications of automation for professional identity, and advance "the need to further study the factors that lead to both the resentment *of* and resignation *to* machines" (p. 39, emphasis in original). As Prahl et al. aptly conclude of the pilots' narrative discourse,

> their actions have spurred a discussion which offers a window into the future of the complicated relationship between human and machine. A decade from now, we are sure to be witnessing these same tensions in the countless other industries implementing automation. At that time, we may not look back to aviation and think 'we've seen this movie before,' but we can rest assured that aviation and the 737 MAX incidents provide a sneak preview. (p. 40)

Aviation is a paradigmatic sector for the history of automation. Here the first autopilot was introduced. The recent tragedies of air travel teach that the development of automation takes place in a world apart from the concrete practices of daily organizational life and if its outcome is not sufficiently communicated and explained, it cannot be integrated within the world of aviation effectively. From Prahl et al.'s research emerges clear evidence that the division of labor between machines and pilots has not been sufficiently outlined; consequently, neither has the division of roles between them. Therefore, the autonomy of the machines takes the power of decisions from pilots to elsewhere. Problematic management of human resources also emerges: aviation companies focus more on machines than on human pilots. This means that human resources management in this sector has become a sort of peripheral objective rather than the core of the related business model. In addition to being dangerous for both pilots and passengers, this strategy is self-destructive on the part of companies, because if they do not understand that they have to correct the trend the sector risks crashing. As Darling (2021) pointed out, "when talking about robots, anthropologists Alexandra Mateescu and Madeline Clare Elish like to use the term 'integrate' instead of the more commonly used word 'deploy' because, as Elish says, 'integrate' prompts the question 'into what?'" (p. 49). Relationships among automation and labor require holistic consideration of the complex contexts, systems, and structures in which they are formed.

Prahl et al.'s paper also leads us to reflect on the current ideology that drives managers not only in the aviation sector but also, for example, in the automotive sector and it highlights the immediate necessity to explore their opinions, attitudes, and behaviors on an empirical level. What do they imagine about machines, automation, and artificial intelligence? What do they think of the workers' value and role? How do they conceptualize workers: Are they viewed as error-prone remnants of the past or do managers think that it is economically advantageous to reduce the work they have to do? And what is (and should be) the role of regulatory bodies in the face of current developments in automation and artificial intelligence?

In the second article, Riley Richards and Associate Editors Patric Spence and Chad Edwards offer a broad perspective on research trends and trajectories in HMC. In "Human-Machine Communication Scholarship Trends: An Examination of Research From 2011 to 2021 in Communication Journals," Richards et al. present the results of a content analysis of articles published over the last decade in 28 communication-specific journals (9,000+ articles). The results paint a valuable picture of the incidence and influence of HMC scholarship relative to the larger scholarship in Communication Science/Studies. This is useful to understand the state-of-the-subdiscipline and identify areas of strength as well as gaps in our collective research endeavors. Richards et al. shed light on how HMC is being constructed through scholarly publication by determining the percentage of regional, national, and international communication journal articles centering HMC, and codifying the contexts of study and methodological approaches. As they report, only a small fraction of communication research pertains to HMC, but since 2018, 2 years before the establishment of the *Human-Machine Communication* journal, there has been rapid growth that "highlights both the need for the journal and the emergence of the field of study" (p. 52). Their analysis reveals that the most examined context of HMC concerns interpersonal communication/relationships, which is in line with a pervasive tendency among HMC scholars to approach machines as interpersonal or quasi-interpersonal others in communication

(Fortunati & Edwards, 2020), and that the most examined specific settings of interaction include VR/AR and HRI. They report that HMC scholarship reflects a mix of methodological approaches and an interdisciplinary character. Richards et al. conclude by highlighting the need for more qualitative, rhetorical, observational, critical, and theory-building work.

This content analysis brings us back to the useful exercise of questioning how the field of HMC research moves and where it goes. Researchers today experience a great tension between the need for specialization and the desire to keep an overview of the great analytical frameworks in which the issues they investigate are situated. This tension is difficult to handle because scholars are forced to mediate between these two divergent tendencies (Cacciari & Franck, 1981). The solution invoked by Richards et al. also is to maintain a strategy of inclusiveness. This strategy is of vital importance to framing the psycho-social and cultural aspects of HMC within society. Societies have become increasingly complex systems in which the potential increase in social relationships (offline and online) is accompanied by an unprecedented speed of change. In the last few decades, social sciences— understood as systems of knowledge, theories, and methodologies, which serves to explain the social world and its changes—have had to face at least four major disruptions that have had a notable impact on the world of knowledge:

(1) The globalization and internationalization of knowledge, research networks, and the scientific community, which on the one hand has enormously expanded the amount of scholarship produced in terms of books and articles (e.g., Knight, 2007) and on the other hand has intensified the competition between scholars to achieve status in a knowledge-production enterprise measured with bibliometric instruments (e.g., impact factor) (see Gingras, 2014).

(2) The advent of the internet, which has simultaneously offered a series of very useful tools (Wikipedia, search engines, automatic translators, automatic correctors, etc.) and built a parallel and intersecting reality with the offline one, where the production of knowledge from below is confused with multiple manipulation strategies from above (e.g., Elwood, 2008; Gläser, 2003).

(3) The lack of new theoretical frameworks and methods, since, in practical terms, we are using century-old methodological tools, excepting some new applications such as structural equation models, multilevel linear regression models, sentiment analysis, and text mining. We face, as Richards et al. point out, a dramatic lack of appropriate and updated theories to understand the meaning of machines in communication.

(4) The crisis of the university as a place traditionally dedicated to the production of knowledge but subject to the development of centripetal forces of dissolution of knowledge itself (on the one hand, difficulties in staying up-to-date given the sheer number of books and articles that should be read and difficulties in understanding society holistically due to disciplinary barriers that make fields of investigation increasingly narrow; on the other hand, students who are insufficiently skilled in literacy, written and oral expression, and critical thinking) (e.g., Christopherson et al., 2014).

Social sciences today are forced to face the challenge of complexity imposed by these four disruptions to continue interpreting society and predicting trends in social phenomena, structures, and dimensions: Communication is the first of these. Complexity does not mean that there is no longer anything simple in society, but that it is more difficult to get to the simple. What strategies have been applied to deal with complexity so far? Roughly, there

are three: (1) the development of specializations and fine-detailed analyses, (2) the shift in focus from causation to associations, and (3) the use of interdisciplinarity. These strategies have come with some advantages and some disadvantages, but because the disadvantages have often outweighed, we believe it is appropriate to consider alternative strategies for dealing with the complexity. For instance, one technique may be to resort to the so-called Pareto principle (or the 80-20 rule) to identify which are the most relevant factors (20%) to interpret for the current society. Another insight that Richards et al. give us is the need for building HMC theory. While we hope that this article will be an important stimulus for early and more established researchers to advance new theories, it is also crucial to work on our research questions. Framing the psycho-social and cultural aspects of HMC in a post-COVID era requires constructing a different picture of the relationship between society and technology. Humankind has experienced a new centrality of machines in their lives on a massive scale. To work or to teach/learn over the past few years has required extensive use of computers and mobile phones. What has all of this meant in terms of our cultural attitudes toward machines? What were the most relevant changes? Have we become friends with machines or do we feel subjugated? Or, to ponder from the more macro level, what do we want from machines?

Perhaps *meliorism*, in the unique sense it was figured in pragmatist social theory by William James (1977), is a useful perspective to dust off and resurrect at this juncture because it is the doctrine comfortable resting on the brink of both hope and despair for the future while emphasizing that the realization of specific futures depends on the choices we make in the present. Meliorism stands outside the deterministic binary of naïve optimism and cynical pessimism about the course of humanity to underline that the world can be made better by human effort: "It holds up improvement as at least possible; whereas determinism assures us that our whole notion of possibility is born of human ignorance and that the necessity and impossibility between them rule the destinies of the world" (p. 54). In the meliorist spirit, the what-if becomes as or more important than the what-is, and the task of social theorists and social life more generally is to link choice and consequence toward the realization of ameliorative futures (James, 1907/1991). Realizing the potentials of this new field of HMC requires being proactive rather than merely reactive to past and present developments. There is a greater chance for good if we are poised and ready to play a shaping role in the design, implementation, regulation, and socialization surrounding human-machine communication. This requires taking stock, based on the best available thinking to date, of *what we believe HMC needs to be and not to be, whether we are heading in the right direction, and which actions will be most beneficial.*

There is a productive tension generated by the field's need to simultaneously study what is and anticipate what if. An anticipatory positionality recognizes the various ways in which historical and current conditions may, through human action and technological capabilities, lead to potential futures that benefit or burden our identities, ideation, relationships, cultures, and social structures to varying degrees of desirability. Staying abreast in our research and criticism often necessitates the use of anticipatory methodologies, including experiments or surveys which introduce plausible HMC partners, communication practices, and contexts of interaction that are not (yet) pervasive in naturalistic settings, as well as the use of anticipatory theoretical lenses which seek to address the possible alongside the actual of HMC. Experiments, in particular, have emerged as a useful design for the field's

anticipatory posture (Richards et al., this volume). Often praised for high levels of control and maligned for contrived situations, experiments nonetheless offer a singular advantage in modeling possible futures by allowing the creation of bounded social realities reflecting communication conditions that may or may not come to pass, but which cannot (yet) be observed. Thus, there is always a tradeoff: experimental conditions are simplified and unrealistic, but they are also a rare means of generating data and serendipitous findings on simulated speculative futures of HMC. The key is that anticipatory designs must be informed and accountable to the teachings of the long history of communication and technology and naturalistic inquiry of HMC in the present (e.g., Prahl et al.).

Asking the Big Questions: The Other and the Hybrid in HMC

The next articles in the volume contribute in exactly those ways encouraged by Richards et al. (2022) because they feature HMC inquiry at the philosophical, theoretical, critical, and synthetic literature review levels. David Gunkel, in his essay "The Symptom of Ethics: Rethinking Ethics in the Face of the Machine," offers an elegant and provoking application of "the symptom," figured by Žižek as "that excluded 'part that has no part'" (p. 68) to characterize the place of the machine in ethics. The essay progresses by first tracing Žižek's unique operationalization of the symptom, then demonstrating how the machine is the symptom of ethics by its definitional always/already exclusion from the circle of moral philosophy, next considering the available but flawed contemporary attempts to accommodate the machine in ethics, and finally discussing the implications of understanding the machine as the symptom for the excluded other and communication ethics more broadly. "What we need to do in the face of the machine . . . is to recognize the symptom as such and allow it to question the entire history of ethics and its necessary and unavoidable exclusions" concludes Gunkel. "The challenge presented to us by the machine is not just a matter of applied ethics; it invites and entrains us to rethink the entire modus operandi of moral philosophy all the way down" (p. 80).

Gunkel's essay offers an interesting integration of a philosophical perspective into the current debate about mental models and social representations of the place that humans, animals, and machines occupy in the universe (see, e.g., Banks et al.'s 2021 "Forms and frames: Mind, morality, and trust in robots across prototypical interactions). Such models and representations form the basis of people's perceptions, attitudes, and opinions about interactions with artificial agents. This paper provokes the invitation to further explore this theme from a psychological and sociological perspective. The HMC community has already investigated and discussed contemporary ontological boundaries between humans, animals, and machines at a qualitative level (Edwards, 2018; Etzrodt & Engesser, 2021; Guzman, 2020), but there is also the need to go for representative surveys capable of capturing whether the ontological frameworks that affect people's attitudes and behaviors are changing and, if so, in which directions. As a scientific community, we should learn to live with the symptom of which Gunkel talks in his essay and to cultivate it, to understand the strategies with which individuals, groups, and societies cope with the permeation of machines into the social body. Not only will this help us make sense of the innovations and technologies which have shaped our daily lives, but also, in turn, to remodel those technologies

according to our needs and desires (Oudshoorn & Pinch, 2003); not only to give them meaning within other meanings but also to question ourselves about their social meanings and roles.

If a scholarly community asks itself small questions, it will always get small answers that satisfy the citadel of specializations but that leave the city of the general discourse completely unguarded. We must be daring with our research questions—even if we already know from the outset that it is difficult to find a convincing answer—because we are aware of the symptom, that is, of the "part that has no part." We need to continue to explore how laypeople structure the world around them because the changes in the ontological order occurring on social and political levels then go on to enable or constrain transformations of the ethical and philosophical perspectives on rights. We also know that the history of humanity is marked by changes to the structure of the model of the universe. We need to better understand the evolution and history of human culture and to further explore the process of ontogenetic formation of the instability of the borders separating and linking the various spheres of nature, to understand the mental processes through which the tensions between these spheres develop and dissolve, to understand how ontological boundaries are conditioned by a series of historical and cultural sedimentations, and to examine how dreams and imagination relate to these tensions. Moreover, we must also reflect on how visual media—from television onward—have challenged the boundary between reality and its representation.

In effect, if the philosophy of today deals with the theme of accommodating the machine in ethics, the sociology of culture for at least a quarter of a century has been studying the commercial explosion of childhood cartoons, television series, films, and toys that have "the other" as a theme, from robots to hybrids (Fortunati, 1995). We argue that children's consumption of these cultural artifacts worked as a precognizant strategy of what would happen a few years later in factories and the domestic sphere. The periodic return of the collective imagination to strange creatures arising from weakening ontological boundaries has been possible because these creatures have always been part of human beings' cultural history.

In the fourth article, "Human, Hybrid, or Machine? Exploring the Trustworthiness of Voice-Based Assistants," Lisa Weidmüller investigates the applicability of predominant models of trustworthiness to VBAs, which may be considered *hybrid* communication technologies in the sense they are often perceived to be more/other than simply machine. Historically, the trustworthiness of humans and machines has been conceptualized and operationalized distinctly. Whereas human-centered definitions of trustworthiness high-light dimensions of integrity, competence, and benevolence (or character, competence, and caring), machine-centered models stress reliability, functionality, and helpfulness. This opens a question as to which of these approaches to assessing trustworthiness (human, machine, or hybrid) best applies to the emergent ontology of "personified things" (Etzrodt & Engesser, 2021). Results of an online survey of German university students ($N = 853$) and staff ($N = 435$) demonstrated acceptable model fit for both human and hybrid trustworthiness models, but insufficient fit for the machine model; further, fit was moderated by prior experience with VBAs. As Weidmüller points out, this exploratory investigation draws attention to the important topic of valid and reliable instrumentation for measuring HMC

variables in ways suitable to the unique features of context. There are broad implications for the community to test to the extent to which variables, concepts, and constructs originally developed to investigate impressions of humans or machines will fit the human-machine mold, and to explore conceptual and operational hybridity.

Weidmüller's article and the emergent ontology of "personified things" studied by Etzrodt and Engesser (2021) reopen the specific question of hybridity between humans and machines. Voice-based virtual assistants represent the resurgence of the hybrid within the field of social robotics. As we said above, hybridity is not a new issue or a fruit of our contemporaneity. Rather, today's hybrid assemblages may be viewed as modern-day manifestations of ancient and original tendencies. Ancient Greek myths tell us that the boundaries between the various elements of nature were perceived as fluid; the Greeks originated a wonderful repository of hybrid creatures, which demonstrates the collapse of all kinds of borders between beings. Even the boundary between deities and humans was malleable to the point that their unions gave life to the demigods, or the half-human/half-gods, who were endowed with superhuman powers but were not fully divine. In fact, demigods were usually mortal, except in some very rare cases (e.g., Dionysus).

If the boundaries between deities and humans were permeable, even more so were those between humans and the world of animals, plants, and minerals. Let us start with the first family of hybrids, which consisted of fusions between humans and animals, and was perhaps inspired by fears of "unnatural" relations between the two (e.g., bestiality practices). To recall a few, there were the Harpies (women's head, vulture's body, dragon wings and claws, bear ears, and wolf teeth), the Sirens (bird-women with florid breasts, feathered wings, feminine face sometimes bearded, rapacious claws, lion's paws, and egg-shaped lower body), the Centaur (half man and half horse), and the Satyr (man's body with ears, tail, and possibly horse or goat hooves) (Gigante Lazara, 1986, p. 11). The second family is composed of the hybrids between humans and plants, and includes, for example, the Botuan, a man-plant with a human face, arms and feet, and a palm body, and the Wak-Wak, a mythical tree whose fruits, which were human beings, detached and fell to the ground when ripe. This tree, which probably represents a version of the Tree of Life, was present in numerous cultures with variations: in the Chinese version, children were born, in the Indian version girls, in the Arab version unidentified living beings, and in the European version women, or the tree itself was the transformation of human beings (Baltrušaitis, 1982, p. 130; Dal Lago, 1991, p. 228).

The third family included the human/stone hybrid. In many cultures, it is told that humanity was born from stones. It is also handed down that the Persian god Mithra was born from a stone; in Phrygia, the Great Mother Cybele was a stone fallen from heaven; according to the myth of Deucalion and Pyrrha, they became the progenitors of a new human race, since they threw stones behind them from which the new humanity was formed. The age-old personification and deification of stones explain why in ancient Rome, some altars and statues were reported to sweat, bleed, or even shed tears (Bloch, 1981, p. 101), a phenomenon that continues up to the present day.

The question of hybridization not only involved fusions between humans and other elements of nature, but also fusions between other elements of nature with each other. The fourth family included hybrids between animals and plants, such as pomegranates that

produced birds, or branches that, once fallen from the tree, dragged themselves along like snakes (Baltrušaitis, 1982, p. 131), as well as hybrids between animals and minerals, such as the zoomorphic depiction of nature as a whole (jaws, beaks, and faces were outlined on nature, which appeared threatening in every aspect; Baltrušaitis, 1982, p. 214). The idea of zoomorphic nature originated in the East—in the Chinese topographical system Feng-Shui, the earth's crust was seen as traversed by secret forces composed of a male matrix that corresponded to the blue dragon and a female matrix that corresponded to the white tiger—then passed through Egypt and arrived through Greece in the West. In the Renaissance, for example, Leonardo da Vinci suggested applying the same method he developed to observe and interpret the stains that formed on the walls to the reading of animal forms in nature. An analogous conception of the earth-animal is found in Mexico, where an insatiable toad-shaped monster devours not only the dead but also the sun and the stars at sunset (Neumann, 1981, p. 185).

The family of hybrids most interesting for our discourse is that including the hybridization between humans and inanimate objects. This innovation traces to Hieronymus Bosch, who painted objects in union with quadrupeds as well as human beings, such as vases equipped with a woman's bust and a donkey's head. In this integration of objects, it was once again the East that gave a rich life to the union of tools and human beings. The oldest depictions of object processions come from Japan in the form of an overturned bowl that runs on its handles, a suitcase with eyes and a mouth for a lock, and a sheathed knife that trots on two legs. Japan also had personified objects: these were kitchen utensils and humanized boxes with human form and intelligence. This trend continues up to the present day and manifests, for example, in the alphabets in which the letters are formed by human characters (Muratova, 1985, pp. 1359–1360). In this framework, the hybridization between humans and machines deserves special comment. The figure which probably embodies this phenomenon most directly is Talos, a being made of bronze, half-human/half-automaton, whose task was to protect Crete (Magnenat Thalmann, 2022). Because of its metallic nature, it could jump into the fire to become hot and then pursue enemies forcing them into a deadly embrace. Talos was kept alive by a single vein that crossed its body from neck to ankle, where it was closed by a nail or membrane to prevent spillage of its vital liquid.

Thus, at the level of culture, Talos's myth tells us that the hybridization of humans with machines is one of the most prominent archetypes of humanity. Arguably, the hybrid is such a vital cultural component of the archeology of imagination that it lives in every child. On this idea, Freud (1990) maintained that "the child is forced to recapitulate during the early stages of his [sic] development all the changes in the human race" (p. 234). Expanding Freud's intuition, Piaget and Inhelder (1970) stated that

> the child explains human beings to the same extent that they explain the child, and often more, since if the first educates the second through multiple social transmissions, every adult, even if a creator nevertheless began with being a child. (p. 9)

In the same vein, Lévi-Strauss (1958) affirmed that the psychology of the very young child constitutes "the universal fund infinitely richer than that available to any particular society." And again: "at birth, and in the form of sketched mental structures, every child bears the

entirety of the means which humanity has at its disposal from eternity to define its relations with the world" (pp. 119–120).

Other processes and elements, however, resonate with the hybrid and contribute to further shaping how humans experience it. They are the development of the psychology of the child and the various stages of the construction of reality as well as the formation and structure of the imagination. Symbolic thinking together with the original ghosts takes shape only from a certain point in the child's life which is, precisely, in the proto-perceptive and proto-cognitive space. The newborn, at the moment of birth and in the earliest period of extra-uterine life, is scarcely able to differentiate the perceptive capacities addressed to the external world. Newborns are unable to perceive their organism as separate from what surrounds them, and thus they are unable to focus upon and distinguish the nuanced characteristics of external objects. In this context arises the experience of trespassing boundaries and the possible interchange and sliding of some characteristics from one object to another. This is where infants relive the dimension of the hybrid and, once experienced, this dimension will later claim its presence within the fully-developed structures of the individual, contributing to the constitution of dreams, daydream fantasies, and creative processes (Funari, 1988, pp. 27–29). In dreams, also, the experience of the hybrid presents itself as the phenomenon of condensation, which is the fusion of two images into a single composite image; for example, a person who unites the features of two distinct people (Piaget, 1982, p. 184). After this type of assimilation, thought cannot undo the union and fully differentiate the two people in a way that returns them to their original features. The composite image remains at least at an unconscious level in the imagination, even if we no longer remember it consciously. Durand offers us a great contribution in understanding how the hybrid is present in the imagination of children and also remains in our imaginations as adults, without there being an ontological first between the two. Indeed, it is impossible to scientifically establish ontogenesis and phylogeny of symbols, and so it is convenient, suggests Durand (1987, p. 29ff), to place oneself on what he defines as an "anthropological journey," a place where there is a two-way reciprocal influence of inner drive and the surrounding material and social environment. At the conclusion of our discourse on the hybrid, we can see that voice-based virtual assistants can be perceived as hybrids because we have a pre-existing and primeval form of this concept.

Addressing the Domestic Sphere: Gender, Labor, and the Political Economy of HMC

The context in which the next three articles are situated is in the domestic sphere. HMC is analyzed at various levels in relation to communication, sexuality, and care. The domestic sphere is the dominant sphere of the capital system and is where the highest amount of value is extracted by the five giants of the web: Google (Alphabet), Apple, Facebook (Meta), Amazon, and Microsoft. These monopolistic multinationals have rapidly expanded especially in the West, creating a techno-information complex that has created an impressive capitalization, further facilitated by tax avoidance and political lobbying. "The combined yearly revenue of Amazon, Apple, Alphabet, Microsoft and Facebook," writes Shira Ovide in *The New York Times* (October 12, 2021), "is about $1.2 trillion, according to earnings reported this week, more than 25 percent higher than the figure just as the pandemic started

to bite in 2020. In less than a week, those five giants make more in sales than McDonald's does in a year." The domestic sphere, in addition to talking about an enormous extraction of capital value in areas such as communication, sex, and care, talks also about the hybridization that makes it impossible to distinguish the contribution of HMC from mobile communication and from that mediated by computer in the process of value formation. We cannot forget that, although each of these forms of communication contributes to altering in a specific way everything each finds in their paths, which makes it worth it to analyze them separately, they are all fueling the insatiable appetite of digital technology that is among the most powerful economic forces shaping the world today. This implies not only that the contexts in which communication, socialization, sex, and care occur has changed but also that their social meaning has changed since they have become fields of direct extraction of value in addition to their historical function as areas of domestic labor and thus of indirect extraction of value through the exploitation of the labor force (Fortunati, 2018).

In the fifth article, "Considering the Context to Build Theory in HCI, HRI, and HMC: Explicating Differences in Processes of Communication and Socialization With Social Technologies," Andrew Gambino and Bingjie Liu make a significant contribution to theory-building by demonstrating specific ways in which digital HMC processes may differ from interpersonal, face-to-face processes. In the first volume of *Human-Machine Communication,* Gambino et al. (2020) proposed an extension of CASA, stating that people do not necessarily apply social scripts associated with human-human interactions as claimed by CASA theory, but perhaps also social scripts associated with interactions specific to media entities. In the present paper, Gambino and Liu build their theoretical contribution via a comparison perspective, which has a long tradition in the history of communication studies as means to identify both the differences and similarities, the advantages and disadvantages of the various forms of in-person and mediated communication.

As Gambino and Liu point out, one of the challenges of theorizing HMC is that our research cannot always match the pace of technological development, which has necessitated a focus on affordances, features, and use practices with probabilistic effects across many media, technologies, and platforms (p. 112). For example, Carmina Rodríguez-Hidalgo, in the inaugural (2020) volume of *HMC* introduced a model of enacted communicative affordances which reconciles the robotics and communication science perspectives in light of unique aspects of communication with social robots. Gambino and Liu propose a brilliant and broadly-useful avenue for HMC theory-building rooted in "consideration of the relationship between *contextual* factors in HMC and those in theories of communication and relationships" (p. 112). The heart of the essay is an elaborated demonstration of how two existing theories of socialization and message production may play out differently in HMC and human interaction because of contextual distinctions. Wisely, Gambino and Liu chose to look at communication as being tightly interconnected with socialization. The linkage between communication and sociality is unavoidable in the practices of everyday life. Take work, for example, or primary socialization processes: How would it be possible to manage, organize, and carry out work without communicating or accompanying children into society without teaching them communication skills? Moreover, communication is not only needed to elicit sociability, but sociability also promotes communication. As Fortunati and Taipale (2012) argued, "since communication is an action with a low output of energy, to be effective and to last over time it has to be embedded in social relations and activities,

which imply a higher energy requirement" (p. 34). What Gambino and Liu show us is the important effect of the circularity of communicative behaviors from one context to another one. For example, they explain how the communication mode with which people approach a machine may reverberate on their interpersonal communication practices, potentially to the detriment of those involved.

The authors also suggest there may be reduced opportunities for observational learning (Social Learning Theory) in the context of digital HMC because it often involves private use. This may, among other things, "lead to a developmental calculus (i.e., the ratio of experiential and observational learning) that relies more heavily on experiential learning" (p. 115). There are massive possible consequences of the proliferation of socialization practices that rest on people's direct experience with machine interlocutors and these include altered behavioral norms and interaction scripts. As Berger (2005) wrote about computer-mediated communication, "the interaction procedures and conventions associated with the use of these technologies may subtly insinuate themselves over time into the conduct of nonmediated social interaction, thus altering the fundamental nature of face-to-face communication" (p. 435). The concern for potential interactions between people's conduct in HMC and their conduct in human social interaction demands greater theoretical and empirical attention. Second, in the case of the goal structure undergirding message production, Gambino and Liu consider differences in the objectives for communication people may bring to their interactions with machines in comparison to other humans, differences arising from context-linked aspects of the (perceived) nature of humans and machines and their capabilities, roles, and functions. For example, face concerns and social judgment fears may figure less prominently in the goals driving message production in digital HMC than human communication. This article by Gambino and Liu is a must-read, not only for its erudite recommendations for advancing HMC theory but for the exemplary exercise of tracing out in two concrete examples those aspects of context that necessitate theoretical reconfiguration.

In the sixth article, "Sex With Robots and Human-Machine Sexualities: Encounters Between Human-Machine Communication and Sexuality Studies," Marco Dehnert examines the topic of sex robots as fertile ground for theorizing from an HMC perspective. Like Gambino and Liu, Dehnert foregrounds *context* as key to theory-building. Further developing a critical approach to HMC (see Dehnert, 2021; Dehnert & Leach, 2021; Fortunati & Edwards, 2020), Dehnert draws HMC and sexuality studies (SeS) into conversation with each other to consider the meanings of intimacy, love, and sexuality among humans and machines. As in the case of ethics and moral philosophy (Gunkel), the machine is also a radical communicative other in the context of sexuality. By intersecting HMC and SeS frameworks, Dehnert pursues a fluid, more-than-human, and ecological conceptualization of communicative sexuotechnical assemblages. There are provocative implications, as Dehnert argues, for both theory and practice in HMC. In terms of theory, these approaches allow for more nuanced perspectives of sex robots that avoid both utopian and dystopian visions of them. In terms of design, Dehnert argues that we must problematize the meaning and representation of sex, gender, age, ability, power relations, and anthropomorphism as they are modeled in sex robots, which must always and only be understood in the context of larger systems of meaning (on this point, see also "Social robots as the bride? Understanding the construction of gender in a Japanese social robot product" by Jindong Liu, 2021).

This paper suggests the need to reflect on the topic of sex robots from a wider perspective. Dehnert has the merit to distill this theme to a concentrated technical analysis, searching along the way for suggestions and perspectives in the literature of sexuality studies. Beyond the conclusions he arrives at, it remains to be understood why human sexuality has been so readily subjected first to the forces of digitalization and then to those of automation. Sexuality is a crucial task, which on the one hand, constitutes part of the unpaid domestic work within the process of the reproduction of the workforce and, on the other hand, encompasses sex work, which may be paid more or less but with the attendant cost of a strong, social stigma. By digitizing sex (making it virtual and disembodied) and now automating sex (making it nonhuman and asocial), what do these forces of mechanization aim to achieve? No doubt, the purpose is to extract more value, as stated above, because part of domestic sexuality work has been transferred online (monetized and compensated) where it is now consumed especially by men, but also by cultivating an ever-larger separation between one individual and another, creating stronger dividing walls between human beings. After the advent of the feminist movement and its various waves, sexuality has become a great field of resistance and struggle on the part of women and LGBTQ communities who have challenged and loosened the coupling between sexuality and the reproduction of children (e.g., Arrow et al., 2021). The regulation of relations between men and women has also been transformed. It is on the wave of these developments that so-called online porn has developed, largely in response to the struggles of women who, for example, no longer enact sexuality in the domestic space to respond only to male needs as they were once historically and socially shaped to do. The specific consequences of the introduction of digital porn and now of sex robots for the quality, value, and meaning of individuals' sexuality must be foremost concerns. The point is that for the capital system not only does sexuality become an additional terrain in which to make money at the expense of consumers but it also becomes a formidable terrain of control and command over people's sexuality through machines. According to Dubé and Anctil (2021, p. 1206) "the private sector is racing to develop new erotic products to occupy an untapped sextech market that is estimated to be worth $30–120 billion." It is within this broad contextual understanding of commercialized digitization and automation processes that we must further interrogate what value of sexuality is lost, gained, or transformed for individuals.

With the next article, we conclude the trilogy of articles whose topics align with the domestic sphere. In particular, we come to discuss how a specific type of care work in society has been picked up by machines. Austin Beattie and Andrew High, in the sixth article, "I Get by With a Little Help From My Bots: Implications of Machine Agents in the Context of Social Support," present an implication-rich synthetic literature review of HMC research relevant to seeking emotional support from bots. Beattie and High ground their project in the historic and rising contemporary developments in chatbots designed to provide therapy, emotional assistance, and supportive interactions. Based on their organized review of research on seeking and processing support in HMC, Beattie and High articulate conflicting and testable perspectives touching the heart of an important issue for both researchers and users of such technologies: When one partner is a bot, are the processes and outcomes of social support richer (*improvement perspective*) or poorer (*impairment perspective*)? Resonant with the theme of *context* threading through this volume, Beattie and High suggest that as a starting point for further inquiry, researchers should attend to "the characteristics

and qualities of contexts in which machines may impair or improve supportive outcomes, as well as how factors such as technological efficacy, the severity of the stressor, how stigmatizing a stressor is perceived to be, or a bot's degree of humanness influence the process of support represent several clear starting points for further inquiry" (p. 162).

This synthetic literature review invites us to reflect on what has happened and is currently happening to care work in society. Why do people feel compelled to or choose to turn to machines for emotional support? Once again, the answer to this question is to be found at least in part in restructured gender and labor relations, since care and support work of this nature has historically been associated with women and the domestic sphere. On the one hand, women face the impossibility of being the sole performers of family care due to the overall rhythms of their work, which often form a non-stop continuum between housework and waged work; in this sense, care work and that demanded by the market have become irreconcilable because of practical demands on time and energy. On the other hand, they may face the desire to put an end to an unpaid job falling mostly to them within families as a reaffirmation of their right to social appreciation and the economic regulation of domestic work. People of advanced age have been the first to experience the consequences of the impossibility for families (and especially women) to perform the entirety of care for their members. In the redistribution of domestic work within couples that took place after feminism, men took on only a part (varying from country to country) of this work. At the same time, governments of many Western countries (including the US during the Trump administration) have made substantial cuts to their social welfare systems (The Lancet Commissions, 2021), reducing or eliminating adequate funding for nursing homes, in-home eldercare, and paid family leave to care for older family members, as well as social services for children such as nurseries, kindergartens, after-school activities, cafeterias, and dedicated public transportation. The potentially dehumanizing aspect of this social and economic organization is that it both strips people of the time and resources needed to take care of each other, while simultaneously throwing more care work back onto the shoulders of individuals. Technologies can help in caring (because a robot or a bot is likely better than having no one), but automated care may also risk a relevant deterioration of the quality of care work. In their article, Beattie and High foreground this important possibility and offer paths for investigating when, why, and how the use of bots may improve or impair care.

Examining Industry: HMC Applications and Theory in Education and the Factory

The last two articles in this volume deal with specific and future-facing social contexts involving the use of AI and robotic technologies: education and the factory. AI and robotics arrive in societies which Bolin (2022) describes as

> late, modern life in the technologically advanced parts of the world can be described as having entered a deepened form of informational capitalism, based on the datafication of social domains and colonizing parts of our lifeworlds that have previously been outside of the market (Bolin, 2018; Couldry & Mejias, 2019; Snircek, 2017). (p. 23)

Neo-capitalism affects all societal domains and these final articles examine two important sectors—education and the factory—which are undergoing particularly rapid and dramatic automation processes at this moment. Let us start with education, which belongs to the sphere of social reproduction and is the place where knowledge is shared with young generations to prepare and train them for their future work and professions, and to socialize them for public life. Over the past few decades, this sector has been colonized by processes such as computerization and datafication (van Dijck et al., 2018) and because of the COVID-19 pandemic, it has been forced to transform its main tasks—teaching and learning—from in-person to online or hybrid activities. Education stands out as especially affected by the recent wave of neo-capitalism because the opportunities for platform companies to capture new market shares have rapidly multiplied. Looking at the school sector, the OECD (2021) writes:

> Last year, 1.5 billion students in 188 countries were locked out of their school. Some of them were able to find their way around closed school doors, through alternative learning opportunities, well supported by their parents and teachers. However, many remained shut out when their school shut down, particularly those from the most marginalized groups, who did not have access to digital learning resources or lacked the support or motivation to learn on their own. The learning losses that follow from school closures could throw long shadows over the economic well-being of individuals and nations. The crisis has exposed the many inadequacies and inequities in our school systems. (p. 3)

Even at the higher education level, which is presumed less vulnerable than primary and secondary schools because university students are expected to be self-directed and independent in their studies, this pandemic has shown that aside from questions about the efficiency of online teaching and learning, the move from teachers' autonomy to automated data analytics (van Dijck et al., 2018) has been tumultuous. At the same time, COVID-19 has shown that it was cost-saving to offload the expenses for rooms, energy, IT support, Wi-Fi, and equipment onto teachers and students, to the extent that some worry there is a risk that "lecturers and researchers will become freelancers or subcontractors in the trade of knowledge" (Bolin, 2022, p. 31).

It is in the light of this framework that the eighth article, "Embracing AI-Based Education: Perceived Social Presence of Human Teachers and Expectations About Machine Teachers in Online Education," by Jihyun Kim, Kelly Merrill Jr., Kun Xu, and Deanna Sellnow should be read. In particular, this article reports the results of an online survey exploring whether and how students' prior experiences with human-taught online courses were linked to their expectations of AI teaching assistants in the future. The results indicated that the social presence (psychological involvement) of the human teacher was associated with more positive attitudes toward an AI teaching assistant and higher intentions to adopt the technology if given the opportunity. Kim et al. explore the meaning and implications of the link between experiences with humans and expectancies of machine partners. "It is not clear yet when a machine teacher or AI-based education will be readily available in higher education" (p. 178) they note, but their findings reinforce and reflect the reality that when

machine partners fulfill roles traditionally performed by human beings, there will be cognitive, affective, and behavioral carryover that colors reactions toward them.

The other social context analyzed by the ninth and final article is factories, which are places with a long history of powerful impacts by machinic development. This article is entitled "Exoskeletons and the Future of Work: Envisioning Power and Control in a Workforce Without Limits" and is written by Gavin Kirkwood, Nan Wilkenfeld, and Norah Dunbar. Kirkwood et al. offer a valuable contribution to theorizing and gaming out the implications of wearable robotics, which present a site for investigating the literal, embodied combination of human and machine in the workplace. Specifically, their purpose is to "discuss the potential of industrial exoskeleton technologies to shape human-machine and human-human power relationships across a variety of industries and theorize how power dynamics might change in these settings" (p. 188). Using Dunbar's interpersonally-oriented dyadic power theory (DPT) (Dunbar, 2004; Dunbar et al., 2016), Kirkwood et al. artfully trace how power, interpersonal dynamics, and autonomy may be disrupted in the face of emergent technologies. They demonstrate the applicability of DPT to the context of exoskeletons in workplace HMC and offer revised theoretical propositions adapted to the unique contextual features which arise. Issues of authority (who ought to have control) and resources (including their implications for diversity, equity, and inclusion) are of utmost importance in human-machine configurations, and there is much to cross-apply from their negotiation in interpersonal communication. At the same time, the differences to these processes introduced by HMC necessitate theoretical extension and refinement. To that end, Kirkwood et al. offer a series of revised propositions intended to forward an expanded theory useful for studying the integration of exoskeletons in workforces.

Exoskeletons are an interesting topic to investigate and reflect upon since they are the contemporary emblem of an important move that is happening in the factory. As in the domestic sphere and in the sphere of social reproduction, machines have come increasingly close to the human body (e.g., mobile phones, laptops, Google Glass, Oculus, Google watch, and so on) and even penetrate it (e.g., the pacemaker). In factories, also, machines increasingly approach the human body of workers. The consequence is that workers today have shifted from being machine appendages to becoming an integral part of the machines themselves. Exoskeletons are one of the typologies of robotic machines present in today's factories. Traditional factory robots have helped pave the way for collaborative robots or cobots, which, unlike most of the industrial robots adopted up to 2008, which were designed to operate autonomously or with limited guidance and were protected by barriers, are designed to physically interact with humans in a workspace. Collaborative robots, however, along with the cluster of technologies and processes such as soft automation, digitalization, AI, big data, social media, and 3D printers (Evans, 2017), which are redesigning the factory world today, are far from being truly collaborative. Instead, they mainly contribute to dictating the pace of work. Delfanti (2021) for example, has focused on the effects of this technological and organizational regime on Amazon workers, analyzing their struggles across the world. He reports that in their protests against the work rhythms imposed by the machines "to increase their productivity, standardize tasks, facilitate worker turnover, and ultimately gain control over the workforce" (p. 40), their slogan was "We are not robots!"

The article by Kirkwood et al. on exoskeletons is an important addition to the study carried out by Andrew Prahl "Fight for Flight: The Narratives of Human Versus Machine

Following Two Aviation Tragedies" to investigate the critical issues of the recent developments in aviation automation and to the studies published in Volume 2 of this journal by Piercy and Gist-Mackey (2021) on workers' anxiety in respect to the diffusion of automation in the pharmacy context and by Prahl and Van Swol (2021) on the behavioral and psychological effects of replacing humans with robots in the financial sector. We hope this accumulating body of work will encourage other contributions on industrial or commercial sectors.

Conclusion

At this phase in the development of HMC as a field of study, there remains the need for more research on actual practices of use that illuminate meaning-making between humans and machines in naturalistic environments (Fortunati & Edwards, 2020, 2021). At the same time, there is the practical reality that many of the technologies and interaction practices of interest in HMC are undergoing rapid development and change, are still at the prototyping level (Fortunati et al., 2021), or rest on speculative probabilistic futures for communication technologies. Thus, we are faced with dual and sometimes conflicting demands to ground research and theory in the real and unfolding lifeworld of human-machine communication, and to also claim a seat at the table during the earliest phases in design, testing, implementation, law and policy, and ethics to *intervene* for social good. Such interventions must be performed, in each case, before it becomes too late to make a difference in the emerging practical realities of HMC. Taking up this second role requires serious consideration of what we want from HMC, honest assessment of whether we are moving in productive or unproductive directions in society, and strong advocacy for those designs and practices which safeguard the dignity, inclusivity, and well-being of humanity and other earth communities. After all, the main trajectory of social theory passes through the level of description, which enables explanation, which allows prediction (Albridge, 1999), which finally empowers *action*.[1]

References

Aldridge, A. (1999). Prediction in sociology: Prospects for a devalued activity. *Sociological Research Online*, *4*(3), 5762.

Arrow, M., Barrett Meyering, I., & Robinson, S. (2021). In the wake of the sexual revolution: New histories of sexual and gender politics in contemporary Australia. *Australian Historical Studies 52*(3), 313–316.

Baltrušaitis, J. (1982). *Il medioevo fantastico*. Mondadori.

Banks, J., Koban, K., & Chauveau, P. de V. (2021). Forms and frames: Mind, morality, and trust in robots across prototypical interactions. *Human-Machine Communication*, *2*, 81–103. https://doi.org/10.30658/hmc.2.4

Berger, C. R. (2005). Interpersonal communication: Theoretical perspectives, future prospects. *Journal of Communication*, *55*(3), 415–447. https://doi.org/10.1093/joc/55.3.415

---

1. Comte, A. (1830). *Cours de philosophie positive par M. Auguste Comte, . . . Tome premier [-sixième]: Les préliminaires généraux et la philosophie mathématique* (Vol. 1). Bachelier, libraire pour les mathematiques.

Bloch, R. (1981). *Prodigi e divinazioni nel mondo antico*. Melita.

Bolin, G. (2018). Media use and the extended commodification of the lifeworld. In P. Bilić, J. Primorac, & B. Valtýsson (Eds.), *Technologies of labour and the politics of contradiction* (pp. 235–253). Palgrave Macmillan.

Bolin, G. (2022). The uberization of higher education. In K. Kppecka-Piech & B. Łódzki (Eds.), *The Covid-19 Pandemic as a Challenge for Media and Communication Studies* (pp. 49–59). Routledge.

Cacciari, M., & Franck, G. (1981). Come governare la "grande città" degli specialismi. *Critica marxista*, *3*, 71–85.

Christopherson, S., Gertler, M., & Gray, M. (2014). Universities in crisis. *Cambridge Journal of Regions, Economy and Society*, *7*, 209–215. https://doi.org/10.1093/cjres/rsu006

Couldry, N., & Mejias, U. (2019). *The costs of connection: How data is colonizing the human life and appropriating it for capitalism*. Stanford University Press.

Dal Lago, B. (1991). *Il sogno della ragione: Unicorni, ippogrifi, basilischi e sirene*. Mondadori.

Darling, K. (2021). *The new breed: What our history with animals reveals about our future with robots*. Henry Holt.

Dehnert, M. (2021). Communication geographies of human-machine understanding: Entangled agencies, synthetic aesthetics, and machine matterings. *Communication Studies*, *72*(6), 1146–1159. https://doi.org/10.1080/10510974.2021.2011360

Dehnert, M., & Leach, R. B. (2021). Becoming human?: Ableism and control in Detroit: Become human and the implications for human-machine communication. *Human-Machine Communication*, *2*, 137–152. https://doi.org/10.30658/hmc.2.7

Delfanti, A. (2021). Machinic dispossession and augmented despotism: Digital work in an Amazon warehouse. *New Media & Society*, *23*(1), 39–55. https://doi.org/10.1177/1461444819891613

Dubé, S., & Anctil, D. (2021). Foundations of erobotics. *International Journal of Social Robotics*, *13*, 1205–1233. https://doi.org/10.1007/s12369-020-00706-0

Dunbar, N. E. (2004). Dyadic power theory: Constructing a communication-based theory of relational power. *Journal of Family Communication*, *4*, 235–248. https://doi.org/10.1080/15267431.2004.9670133

Dunbar, N. E., Lane, B. L., & Abra, G. (2016). Power in close relationships: A dyadic power theory perspective. In J. A. Samp (Ed.), *Communicating interpersonal conflict in close relationships: Contexts, challenges and opportunities* (pp. 75–93). Routledge. https://doi.org/10.4324/9781315774237

Durand, G. (1987). *Le strutture antropologiche dell'immaginario*. Dedalo.

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.

Elwood, S. (2008). Grassroots groups as stakeholders in spatial data infrastructures: Challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, *22*(1), 71–90. https://doi.org/10.1080/13658810701348971

Etzrodt, K., & Engesser, S. (2021). Voice-based agents as personified things: Assimilation and accommodation as equilibration of doubt. *Human-Machine Communication*, *2*, 57–79. https://doi.org/10.30658/hmc.2.3

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

Fortunati, L. (1995). *I mostri nell'immaginario*. Angeli.

Fortunati, L. (2018). Robotization and the domestic sphere. *New Media & Society*, *20*(8), 2673–2690. https://doi.org/10.1177/1461444817729366

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, *1*, 7–18. https://doi.org/10.30658/hmc.1.1

Fortunati, L., & Edwards, A. (2021). Moving ahead with human-machine communication. *Human-Machine Communication*, *2*, 7–28. https://doi.org/10.30658/hmc.2.1

Fortunati, L., Sorrentino, A., Fiorini, L., & Cavallo, F. (2021). The rise of the roboid. *International Journal of Social Robotics*, *13*(6), 1457–1471. https://doi.org/10.1007/s12369-020-00732-y

Fortunati, L., & Taipale, S. (2012). Organization of the social sphere and typology of the residential setting in Europe: How sociability affects the adoption of the mobile phone in rural and urban locations. *Technology in Society*, *34*(1), 33–43. https://doi.org/10.1016/j.techsoc.2011.12.004

Freud, S. (1990). Prefazione a "L'elemento scatologico negli usi, costumi, credenze e abitudini dei popoli" di J. G. Burke (1913) in *Totem e Tabù e altri saggi di antropologia* (pp. 233–235). Newton Compton.

Funari, E. (1988). Introduzione. In J. Laplanche & J. B. Pontalis, *Fantsma originario, fantasma delle origini, fantasma originario* (pp. 9–30). Il Mulino.

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

Gigante Lazara, V. (1986). *Il segreto delle sirene*. Bibliopolis.

Gingras, Y. (2014). *Bibliometrics and research evaluation: Uses and abuses*. The MIT Press.

Gläser, J. (2003). What internet use does and does not change in scientific communities. *Science Communication*, *16*(1), 38–51.

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

James, W. (1907/1991). *Pragmatism*. Prometheus Books.

James, W. (1977). *The writings of William James: A comprehensive edition* (J. J. McDermott, Ed.). University of Chicago Press.

Knight J. (2007). Internationalization: Concepts, complexities and challenges. In J. J. F. Forest & P. G. Altbach (Eds.), *International handbook of higher education* (pp. 207–227). Springer.

The Lancet Commissions. (2021). *Public policy and health in the Trump era*. https://doi.org/10.1016/S0140-6736(20)32545-9

Lévi-Strauss, C. (1958). *Anthropologie structurelle*. Plon (*Antropologia strutturale*, Milano, 1966).

Liu, J. (2021). Social robots as the bride? Understanding the construction of gender in a Japanese social robot product. *Human-Machine Communication*, *2*, 105–120. https://doi.org/10.30658/hmc.2.5

Magnenat Thalmann, N. (2022). Social robots: Their history and what they can do for us. In H. Werthner, E. Prem, E. A. Lee, & C. Ghezzi (Eds.), *Perspectives on digital humanism* (pp. 9–18). Springer.

Muratova, X. (1985). I manoscritti miniati del bestiario medievale: Origine, formazione e sviluppo dei cicli di illustrazioni. I Bestiari miniati in Inghilterra nei secoli XII e XIV. XXXI Settimana di Studio del Centro Italiano di Studi sull'Alto Medioevo, 7–13 aprile 1983, Spoleto, vol. II, pp. 1319–1362.

Neumann, E. (1981). *La Grande Madre*. Astrolabio.

OECD. (2021). *The state of school education: One year into the COVID pandemic*. OECD Publishing. https://doi.org/10.1787/201dde84-en

Oudshoorn, N., & Pinch, T. (Eds.). (2003). *How users matter: The co-construction of users and technology*. MIT Press.

Ovide, S. (2021, October 12). A perfect positive storm: Bonkers dollars for Big Tech. *The New York Times*. https://www.nytimes.com/2021/04/29/technology/big-tech-pandemic-economy.html

Piaget, J. (1982). *Giudizio e ragionamento nel bambino*. La Nuova Italia.

Piaget, J., & Inhelder, B. (1970). *La psicologia del bambino*. Einaudi.

Piercy, C. W., & Gist-Mackey, A. N. (2021). Automation anxieties: Perceptions about technological automation and the future of pharmacy work. *Human-Machine Communication*, *2*, 191–208. https://doi.org/10.30658/hmc.2.10

Prahl, A., & Van Swol, L. (2021). Out with the humans, in with the machines? Investigating the behavioral and psychological effects of replacing human advisors with a machine. *Human-Machine Communication*, *2*, 209–234. https://doi.org/10.30658/hmc.2.11

Rodríguez-Hidalgo, C. (2020). Me and my robot smiled at one another: The process of socially enacted communicative affordance in human-machine communication. *Human-Machine Communication*, *1*, 55–69. https://doi.org/10.30658/hmc.1.4

Snircek, N. (2017). *Platform capitalism*. Pluto Press.

van Dijck, J., Poell, T., & de Wall, M. (2018). *The platform society*. Oxford University Press.

Fight for Flight: The Narratives of Human Versus Machine Following Two Aviation Tragedies

Andrew Prahl[1] , **Rio Kin Ho Leung[2]**, **and Alicia Ning Shan Chua[3]**

1 Wee Kim Wee School of Communication & Information, Nanyang Technological University, Singapore
2 Anglo-Chinese School (Independent), Singapore
3 Nanyang Business School, Nanyang Technological University, Singapore

Abstract

This study provides insight into the relationship between human and machine in the professional aviation community following the 737 MAX accidents. Content analysis was conducted on a discussion forum for professional pilots to identify the major topics emerging in discussion of the accidents. A subsequent narrative analysis reveals dominant arguments of human versus machine as zero-sum, surrender to machines, and an epidemic of mistrust. Results are discussed in the context of current issues in human-machine communication, and we discuss what other quickly automating industries can learn from aviation's experience.

Keywords: human-machine communication, aviation, narrative, automation, qualitative

Introduction

At 37,000 feet over the West Coast of Australia, a First Officer is returning to the flightdeck after a regularly scheduled break in the passenger cabin. As is customary, the Captain asks, "How is it back there?" The expected response is probably something like "good," but a pilot hardly expects to hear "It's carnage out there!" But, that is exactly what occurred on the Qantas Flight 72 (Flight 72) on October 7, 2008 (O'Sullivan, 2017). Just prior, the plane had suffered two sudden dives, all 303 people aboard the aircraft experienced momentary roller

CONTACT Andrew Prahl • Wee Kim Wee School of Communication & Information • Nanyang Technological University • 31 Nanyang Link, Singapore 637718 • andrew.prahl@ntu.edu.sg

coaster-like conditions, from +1g to –0.8g and back again and again. Thrown into ceilings and galleys, 9 of 12 cabin crew members were seriously injured or unconscious. Countless passengers littered the cabin along with broken glass, ceiling panels, oxygen masks, and anything that was not tied down during the harrowing episode. Was it an exceptional bout of turbulence, an accidental control input, a structural failure? Fortunately, quick think-ing by the crew allowed them to identify the offender—the flight control computer—or in other words, the autopilot. The same machine which had controlled over 20,000 hours of safe flight over the plane's lifetime was now violently throwing the plane into a dive. The Captain, attempting to control the aircraft later recounted, "I'm in a knife-fight with this plane. It isn't a fair fight; knife-fights never are" (Sullivan, 2019, p. 2).

Fortunately, the crew of Flight 72 successfully landed the aircraft and all aboard sur-vived. However, just over 10 years later a similar situation ended in tragedy aboard Lion Air Flight 610 (Flight 610). The accident, also involving an automated system that pitched the airplane toward the ground, featured an epic struggle between human and machine lasting nearly 10 minutes (KMKT, 2019). Just months later, a nearly identical accident took place outside Addis Ababa on Ethiopian Airlines Flight 302 (Flight 302). Though Flight 302's pilots had received special training following Flight 610's demise, the crew were still unable to compensate for the commands made by the plane's automated systems. The accidents led to the global grounding of the aircraft involved, the new 737 MAX, which had entered service just over a year earlier (Helmore, 2019).

The resultant accident investigations, legal proceedings, and debates among engineers and pilots are having a profound impact on the aviation profession. Besides the usual sus-pects such as manufacturers, airlines, and regulators, the professional pilot community is also grappling with the meaning of the incidents. Though the debate about automation on the flightdeck has been active for decades prior, the 737 MAX accidents' visceral, physical battle between human and machine is causing long-held animosity toward automation to boil over; but also the uncomfortable acknowledgment by some that the gradual erosion of human authority in favor of machines is necessary. The longstanding debate in aviation is reflective of the tension between professionals and machines in a growing number of indus-tries, including journalism (Carlson, 2015), surgical care (Ruskin et al., 2020) and pharmacy (Piercy & Gist-Mackey, 2021). Machines are also assuming more control in everyday life. From "robo-advisors" managing finances to autonomous cars transporting us, the tension between human and machine agency is becoming more salient each day. As a profession on the frontlines of the human-machine interface, this watershed moment offers an opportu-nity to see how the meaning of concepts like authority, skill, and agency are challenged by machines both in aviation and beyond.

In this paper, we investigate the emergent narratives in the professional aviation com-munity as they process the meaning of the crashes, redefine profession and purpose, and reconcile human and machine on the flightdeck. First, we review aviation's relationship with automation over time and note parallels to key issues in human-machine communica-tion. Next, we briefly summarize the accidents in question and conduct a content-narrative analysis on the world's preeminent professional pilot forum as they discuss the accidents. Results have implications for future debates that will inevitably occur in human-machine communication as machines replace humans.

Automation in Aviation

The image of a pilot struggling with the controls, battling swirling winds, and dodging clouds is no doubt compelling. But such a depiction is closer to Hollywood than the flightdeck of a modern commercial airliner. Stepping into the modern cockpit presents the uninitiated viewer with a kaleidoscope of lights and colorful displays. The lack of *old-fashioned* analogue dials and gauges is a marketing point for manufacturers. The industry-speak for a modern flightdeck presages the dominating role of electronics: they are known as "glass cockpits" (Farvre, 1994). On aircraft from one major manufacturer, the role of the human seems especially diminished; instead of a T-shaped control yoke in front of the pilots' seats, control has been relegated to a small joystick that looks more at home on an Atari than an aircraft.

Piloting an aircraft is a task that lends itself to automation—long, monotonous hours in cruise are simple for a machine to monitor but a monumental task of concentration for humans. Over the course of decades, more automated systems have been added to planes. As such, the broad term *flightdeck automation*—though standard industry jargon—encompasses an ecosystem of various technologies which incorporate everything from robotics to artificial intelligence (AI). For example, while monitoring automation such as a speed-warning system only requires a sensor and display, control automation like the autopilot requires calculation/judgment and robotic control. Recent advances, such as in-flight weather forecasting and automatic route-modification, combine all three—sensors, judgment, and control—with AI. Thus, the modern flightdeck is a quintessential example of the "cluster" of technologies and processes that define modern workplaces where human-machine communication takes place (Fortunati & Edwards, 2021, p. 22). With control automation just now becoming common in production cars, it is indicative of aviation's long history with automation that Tesla's well-known implementation is marketed as *autopilot*.

The increasing capability and presence of automation on the flightdeck has not always been welcomed by pilots. Like other industries such as pharmacy and finance, new technology at times caused *automation anxieties* as the working environment and job scope changed. For example, increasing automation in pharmacy work can lead pharmacists to worry about losing their jobs to machines (Piercy & Gist-Mackey, 2021). This parallels a current controversy in aviation where—citing the capability of automation on the latest jet models—airlines are currently seeking approval for single pilot operation of commercial aircraft (Frost, 2021), a move that has not been received well by pilots and industry groups (Driskill, 2021). In addition to anxieties for aviators, more machines controlling the plane meant pilots felt an increasing "distance" from the aircraft (Tsang & Vidulich, 2002). Several high-profile accidents caused by a pilot's failure to understand complex automation modes, or the inability of pilots to recover when automation failed, led to calls for a new paradigm of "human-centred" automation (Billings, 1996). That proponents chose to call it "human-centred" says both what they wish for and also their view of current technology: indifferent machines designed for an automation first, human second, era.

Agency

The introduction of the Airbus A320 into revenue service in 1988 was a pivotal moment. The new plane was equipped with automatic systems designed to keep the aircraft in the

flight envelope (i.e., controllable state). Unlike other aircraft that would simply warn pilots when flying at the edges of this envelope, the Airbus computers would override the pilots. Airbus also chose to do away with mechanical connections between the controls and the control surfaces—replaced with electronic signals that humans may initiate, but only machines could approve (Farvre, 1994). It was not tacit but outright acknowledgment that ultimate authority over safe flight was best left to a machine; regardless of the truth it was understandably offensive to many pilots (Ibsen, 2009; Sarter & Woods, 1997).

Airbus's decision illustrates how the negotiation of human and machine agency can build new organizational and social structures (Gibbs et al., 2021). In the Airbus structure, humans were symbolically placed in a fenced area, overseen by machines in watchtowers. But, the structure of Boeing's philosophy remained traditional: trust and ultimate control authority remained with the pilots. The rift between the world's two major aircraft manufacturers has endured ever since the A320 flew in 1988, and perhaps the reputation of Boeing as being the *pilot's aircraft* has added to the significance of the 737 MAX incidents studied here. This wasn't *supposed* to happen on a Boeing where human agency is paramount. How could it be that a rogue machine could fight the pilots and win—had Boeing thrown in the towel and silently abandoned their philosophy? With their pride in question, professional aviators anxiously awaited word from investigators. What ultimately emerged tells a complex tale of corporate, regulator, human, and machine failure; we summarize below.

Flight 610 and 302

Both Flight 610 and 302 took off from their respective airports normally in excellent flying weather. But, shortly after take-off and with the plane still at low altitude, both flights experienced sudden pitches downward. On the flightdeck, a cacophony of warnings sounded, including a stall warning (suggesting the plane was flying too slow), and the overspeed warning (suggesting it is flying too fast). Bewildered at this impossible situation, both crews counteracted the pitch down by *pulling up* with the controls. Seconds later, another inexplicable pitch down. Flight 610 continued for several minutes, oscillating up and down as the machine pitched down and pilots pitched up. Aboard Flight 302, the pilots used all their strength to pull on the controls but only managed to maintain roughly level flight for a few minutes. In the end, both flights eventually became uncontrollable and impacted the ground at high speed, leaving only small pieces of debris.

Invisible Machines

The 737 MAX is a modernized version of a much older aircraft that originally entered service in the 1960s. Over many iterations, changes to the length of the plane, fuel capacity, and engine types required minor tweaks on the flightdeck, but by and large the pilot of a 1960s 737 could fly a 2020 version and vice versa. The 737 MAX however incorporated exceptionally large engines that made a substantial change to the way the aircraft handled. In order to spare airlines the cost of a pilot retraining program for a new aircraft, Boeing needed to find a way to comply with regulations stating the new plane must handle like the old one. Boeing's solution was to implement the Maneuvering Characteristics Augmentation System (MCAS), a system that automatically made control inputs to make the plane

feel like previous models. Boeing seemingly did not want many questions about MCAS. Besides ensuring it had a forgettable name and bland acronym, Boeing declined to include information about MCAS in the official pilot training documents for the MAX. And, in the 1,600-page pilot's manual, MCAS was virtually anonymous, mentioned just once.

Silent Malfunctions

The pilots aboard Flight 610 likely had no idea MCAS existed or how to identify its operation. But this was the system that malfunctioned aboard both flights and began thrusting the plane toward the ground. Compounding matters, Boeing had omitted any aural or visual alert of MCAS activation on the flightdeck. And, MCAS had been programmed to move control surfaces faster than the human pilots could (Gates, 2019). The pilots of Flight 302, benefitting from the preliminary investigations of Flight 610, had been briefed on the existence of MCAS, but Boeing had revealed little information about the obscure system and the operational properties of MCAS were still a mystery to the crew on their fateful flight. In sum, the pilots were fighting against a silent machine which was faster, stronger, and ultimately victorious.

Reactions

In the aviation industry, reaction to the crashes has been understandably mixed, with blame placed on everyone from the original designers of the 737 to manufacturers of flight simulators (Helmore, 2019). As more information continues to be revealed in the ongoing investigations, blame has gradually shifted away from the pilots and onto Boeing. Many pilots, having an affinity for Boeing due to their design philosophy, have found this an uncomfortable position to take. In truth, there were a set of actions the pilots could have taken to save the plane, but this sequence is only obvious in hindsight. To gain an understanding of how the professional aviation community is coming to terms with the accidents, we visit the largest online community forum for aviation professionals. Our investigation is driven by three research questions:

> **RQ1:** What are the main points of discussion in the professional aviation community regarding the 737 MAX accidents?

> **RQ2:** How do the 737 MAX crashes challenge or reinforce existing narratives of human and machine in aviation?

> **RQ3:** What narratives emerge from the 737 MAX debate and what do they tell us about the relationship between human and machine in industries beyond aviation?

Methodology

The Professional Pilots Rumour Network (PPRuNe) is a global forum community where professional pilots discuss prominent aviation news. At the time of analysis, the forum

required a registration process that verifies all members are aviation professionals (e.g., pilot, mechanic, cabin crew member). One research assistant gained access to the forum because she was an Air Force pilot trainee. Although the forum requires membership to utilize functions like search and thread bookmarking, all posts on the forum are publicly accessible, indexed by all major search engines, and require no special permission to view. After consulting the Internet Research: Ethical Guidelines 3.0 (Franzke et al., 2020) and the accompanying materials, we determined this to be a low-risk forum (Moreno et al., 2013). Though authors post with no expectation of privacy, and may post behind pseudonyms, we have lightly edited quotes to reduce discoverability. We downloaded discussion threads relevant to the 737 MAX incidents. Upon inspection, the vast majority of discussion took place in the "Ethiopian Airliner Down in Africa" thread. There were 5,124[1] discussion posts from March 10 to May 8, 2019, when the thread was formally closed by moderators. The thread spanned the moment the second crash occurred all the way through the release of preliminary reports from investigation agencies.

Content Analysis

To answer our first research question, we elected to conduct a content analysis of all posts in the thread. Our content analysis method specifies the following steps: (1) initial determination of potential themes, (2) identifying and condensing meaning units, (3) coding process, (4) categorization process (5), relation to emergent themes, and (6) tally count of occurrence frequencies (Erlingsson & Brysiewicz, 2017; Neuendorf, 2002). We performed the content analysis manually using spreadsheet software to inventory and count codes.

Initial Coding Scheme

Our initial coding scheme was developed by reviewing the literature related to aviation and automation. This is an admittedly vast body of research and spans both traditional academic journals and trade journals focused on the aviation industry and human factors. Our initial coding scheme simply served as a guide to guide our initial exploration of the forum postings. After familiarization, our emergent coding scheme took precedence and guided the remainder of our analysis.

Emergent Coding Scheme

Discussion posts were read thoroughly to ensure consistency and comprehensiveness in the identification of meaning units within the thread. Relevant posts were then divided into meaning units which succinctly conveys the essential meaning of the text. Identified units were further condensed while keeping the central meaning intact as demonstrated in Tables 1–3.

---

1. Discrepancies between the number of discussion posts seen in the thread and featured within this research may exist due to deletion by users or restoration of deleted responses after the thread was closed.

TABLE 1   Example—Condensation for #1335
Meaning Unit
"What I want is for the aeroplanes I travel in to be managed safely by pilots of the minimum allowable skill. Not every flight deck will be occupied by superhero pilots"
Condensed Meaning Unit
Aircraft systems should be designed for pilots of low skill

TABLE 2   Example—Condensation for #2396
Meaning Unit
"Reliance on just a single sensor driving a system that was not known and that could take over command is sheer madness"
Condensed Meaning Unit
Faulty design to trust a sensor; Lack of pilot authority

TABLE 3   Example—Condensation for #2894
Meaning Unit
"I actually welcomed the computers to help us to pilot the ships, but increasingly we're only operators and the computers are controlling us"
Condensed Meaning Unit
Automation should assist, but now it replaces; Lack of pilot authority

The next step of the content analysis process is to assign descriptive labels to each condensed meaning unit to assist with the identification of relationships between relevant units as demonstrated in Tables 4–6. Additionally, simple codes were generated to aid in the process. At least two thirds of the posts were checked by two independent coders to ensure agreement. Next, codes were placed into broader categories allowing for the identification of overarching themes.

TABLE 4 Example—Coding for #1335	
Condensed Meaning Unit	
Aircraft systems should be designed for pilots of minimum allowable skill	
Codes	
Low Skill	Simple Design

TABLE 5 Example—Coding for #2396		
Condensed Meaning Unit		
Overreliance affects structural integrity; Pilots should always be in command		
Codes		
Design Flaws	Trust of Machines	Command over Automation

TABLE 6 Example—Coding for #2894	
Condensed Meaning Unit	
Automation should assist but not replace; Lack of pilot authority over automation	
Codes	
Assistance for Aircrew	Command over Automation

Finally, we grouped related codes into categories. We aimed for 10 major categories to clearly summarize the contents of the discussion thread as it relates to our research question. Results are summarized in Table 7 [see appendix for all codes]. Overall, we find that the topics of automation design and human skill dominate the discussion, including how automation should be designed (e.g., redundancy) and what it should communicate (e.g., the human factors of a surprise automation malfunction). Although this is predictable given industry history, we were surprised to see discussions of Boeing's corporate behavior so prevalent. While we would expect topics of corporate culture to be at home on a forum for business managers, its presence here indicates the desire of pilots to understand if the traditional Boeing value of *pilots first* has disappeared. We also see a considerable amount of discussion about what pilots are expected to know and what should be part of pilot training. These topics are clearly related to automation: If modern pilots are no longer trusted to fly, then should pilot training be more akin to an IT professional or should training still focus on manual flying skills? Finally, we note that while human versus machine topics dominated discussion, rarely were both entities talked about in a collaborative frame. In light of the accidents, at least, it was more natural to discuss human and machine working in *opposition* rather than working together.

TABLE 7 Summary Categories			
Number	**Category**	**N**	**Percent**
1	Automation Design: Human-Machine Interaction	458	31.87
2	Human Skill	278	19.34
3	Corporate Silence and Disregard of Safety	256	17.81
4	Automation Design: Hardware and Software Redundancy	138	9.60
5	Regulators and Aircraft Certification Process	91	6.33
6	Human Factors	54	3.75
7	Expectations for Aircrew Knowledge	47	3.27
8	Limitations of Automation and Inappropriate Uses	45	3.13
9	Aircrew Training and Airline Training Philosophy	41	2.85
10	Human-Automation Teaming	29	2.01
	Total	**1437**	**100.0**

Narrative Analysis

Our chosen content analysis method assumes the analysis in step 5 (relation to emergent themes) involves the identification of larger concepts at a higher level of abstraction than content codes. Given the nature of our third research question, we felt it was best to approach this thematic analysis process as a narrative analysis. Although differences between the methods are slight and subjective, narrative analysis is suited to an issue with a long history like humans versus machines in aviation. By "focusing on the *telling* of a story," (Smith, 2016) we look beyond the minutia of technical debates and observe the dominant frames used by pilots to reconcile the events aboard Flight 610 and 302 with personal and professional identity (Black, 2008).

We first identified the posts which were relevant to our research question and excluded posts that were irrelevant, such as news links, technical clarifications, or posts that only included images/emoticons. There were 4,098 discussion posts deemed to be irrelevant, leaving 1,025 posts for narrative analysis. The study team then independently read all relevant posts and met to discuss preliminary narratives. This process was repeated a second time; after the second meeting a research assistant discussed the emergent narratives with two professional pilots with thousands of hours of flying experience for a major international airline. Upon confirmation that our narratives were appropriate, the study team met again to settle on three encompassing narratives.

Human Versus Machine as a Zero-Sum Game

In a zero-sum game, there is no *win-win*. Ideally, automation would be designed in a way that helps human pilots, allowing both parties to leverage their respective strengths to ensure safe operation. For example, humans can use their senses and experience to discern what is the actual state of the aircraft when malfunctioning sensors are sending conflicting messages. Instead, the conflicting inputs on the 737 MAX flights led the machine to essentially pick one as being true and take action. In reality, neither were true. With the proliferation of automation in aviation over the past several decades, one would expect automation designed with human-machine "teaming" in mind (Battiste et al., 2018; Calhoun, 2021). But the narrative emerging from the discussion of the 737 MAX is that the introduction of automation on the flightdeck is associated with *losses* for humans rather than gains. One user comments, "Automation denies skill." Another says, "I think producing pilots capable of hand flying with confidence is an essential skill—it requires a change of culture. Some airlines already do this just by encouraging turning off the automation." The user goes on to recount an interaction with a senior pilot following a tenuous manual landing, " . . . what would the passengers rather have, a perfectly flown approach by the autopilot every time or a pilot who can confidently fly if the situation requires it?" Whereas the first statement suggests physical skills are lost by reliance on automation, both statements acknowledge a psychological skill, *confidence*, is lost as well. One user suggests that automation is preventing the proper training of pilots, "You just need to select and train pilots as it used to be when automation was basic . . . " and another user describes losing touch with the aircraft, "Today, the artificial manual feedback stuff is degrading the primary flight control ergonomics, and distracting and stressing the crew." The commenter starts by saying "today," indicating that

it *used to be* different in the past, like the previous user discussing training. Thus, we see that while new narratives are emerging, the long-running industry narrative about the gradual erosion of human skill is perpetuated in the conversation regarding the 737 MAX.

An important sub-narrative that invites the framing of gains/losses is the characterization of machines as *adversaries*. One user discusses Flight 610, "On [their] last fateful flight, MCAS compounded the crisis by faithfully following its mandated duty—executing a NOSE DOWN infinite loop." The user's anthropomorphising of the machine as a "faithful follower" is attention-grabbing, but the user does not answer the question of who or what issued the "mandate." However, it matters little, the point is not who it was but who it *was not:* the pilots. Another user describes MCAS as one would a maniacal movie villain, "[It is] idiotic that the computers would happily perform a manoeuvre of such violence."

All these excerpts feature a narrative of humans in conflict with machines. Such a narrative is somewhat expected due to the nature of the crashes and the maturity of related narratives in the aviation industry. What is unexpected is the framing of humans losing to machines. And it isn't only skill being lost, machines are causing humans to be stressed, distracted, and lose confidence while the machines are "happy" and "faithful" even as they command the plane toward its doom. One user sums up the narrative, "As an overall improvement we need either reduced reliance on automation enabling greater experience, or improved automation." The user gives no option for experience *and* improved automation, it is *either* humans *or* machines, a zero-sum game.

Surrender to Machines

Another important narrative functions in part as a counter-theme to humans needing to gain control back from the machines. Some pilots eschew the belligerent framing and instead suggest that the battle is already lost. One user asks early in the thread, "Have we passed the point in modern aviation where it is not possible to (quickly) switch off all these pilot and performance aids and fly manually?" Another user poses a question in a whimsical way as to suggest the answer is obvious, but unpleasant, "How does a crew really get back to an old-fashioned 'hands on' configuration of controls, instruments, and sensors (where all of the automation and fancy gadgets are completely isolated from involvement) these days?" Other users suggest that humans are no longer a solution to machines' problems, "Actually I agree with [other user] and concur that the inevitable and unstoppable answer to imperfect automation is improved automation."

Though this narrative is prevalent, it is clearly controversial and virtually every post expressing the view is quickly met with retorts from other users. One user calls his opinion "humble," as if it may be embarrassing to say, "I don't have humble opinions, but the one I do have favours intelligent acceptance of the inevitable AI, not the expensive step in the wrong direction of trying to make new pilots more like the old ones." One wonders about the author's use of "expensive," it appears to carry meaning beyond just money; investing in humans will *cost* safety. Other users are more forthright, "Systems must no longer use human intervention as part of their safety case; we are too unpredictable. Safety critical systems must get smarter; garbage in, garbage out is not an option, neither is giving up and disconnecting."

The narrative of surrender shares similar beginning chapters to the narrative of zero-sum. Both pit mortal versus machine, but the two narratives diverge at the current moment. For zero-sum subscribers, the conflict is escalating and the final chapters are yet to be written. But surrender suggests that the battle is over. One user suggests human skill is itself mortal, "Hand flying is an essential skill. But it is a dying one. The tide, my fellow artful flyers, has turned against us. Were it any other way, we would not be having this discussion." Perhaps the sensitivity of the topic owes to the profound challenge that surrender poses to pilots' professional identity. The symbolism of brass wings and epaulettes is associated with authority and respect, not with mere observers. Nevertheless, some users urge their colleagues toward the realization, the "humble" user quoted above concluded his post with the candid replacement of the word *fly*, "Driving airplanes is not what it used to be, nor should it be."

Epidemic of Distrust

Granting a pilot with control of an aircraft conveys trust from manufacturer, airline, and passenger that the pilot is capable. Aviation has seen numerous episodes involving the distrust (and mistrust) of both humans and machines, but this distrust has typically been limited to specific systems or circumstances. However, the debate about the 737 reveals that distrust has permeated the entire aviation industry. And, though historically supportive of Boeing's stated philosophy placing pilots at the fore, the apparent unravelling of this philosophy leaves pilots dubious toward all industry actors from regulators to airline leadership.

The concept of authority is conceptually related to responsibility; however, we were surprised to see no references to the maxim that pilots have ultimate responsibility for safe operation. Nor was the finger pointed at MCAS alone, instead responsibility was diffused upon many actors. One commenter acknowledged the complexity of the situation but singled out Boeing, "There are many layers of responsibility here, the first being Boeing's insistence on polishing-off a 50+ year-old aircraft . . . bringing questionable stability, and mitigating it all with a poorly thought-through safety system." Another commenter expressed the same sentiment and was doubly sure to spare the individual human engineers and instead implicate management, "I am not sure the blame sits with whoever designed MCAS," because, "I suspect they were backed into a corner constrained by schedule."

Other users reference the Airbus versus Boeing debate, "Boeing went rouge [sic] from their philosophy and instead errantly took a page from Airbus' philosophy, trying to fully automate the plane, and half-assed the entire logic and failed miserably," says one user, "this MCAS system seems criminally designed." Revelations about Boeing's clandestine implementation of MCAS shook pilots, some worrying what else they don't know, "[This all] may have been caused by something else and may be yet another undocumented 'feature' of the MAX." Other users cut into Boeing's press releases, "'To make a safe airplane safer' is just a Coué-method equivalent of 'to make a dangerous airplane a little less dangerous.'" Beyond Boeing, some users have lost faith in the regulatory agencies, "The FAA will undoubtedly approve 'MCAS 2.0', but the evidence shows that there is such a degree of regulator capture that this will be hard to see as an objective evidence-based process." An economic angle is present as well, "In the centuries old battle of profit above all else, safety can only ever come a distant second" laments one user.

In this story of suspicion and doubt, pilots are the clear losers. Machines have no pride to lose, no feelings to hurt, and no concerns of what else may be hidden in the planes they occupy. For some users, fear is palpable. The operation of MCAS is described by various users as "insidious," and "mysterious, ghostly, undocumented." Perhaps envisioning themself in the position of the doomed pilots, a user imagines, "[The crew] were occupied on solving the puzzle and suddenly realized that all this time the ghost in the machine had been busy and still continued incessantly bringing the plane's nose down to the end, it's all too late for them to recover."

Discussion

The 737 MAX incidents offer insight into a profession's struggle to negotiate a shifting identity and purpose as machines play a larger role in the workplace and usurp human agency. For pilots, this workspace is the intimate confines of a commercial aircraft, but this environment is not far different from the hospital operating rooms or car interiors which are being rapidly automated. Unfortunately, the lessons learned from the 737 MAX are the result of a tragic story, but understanding how the professionals on the front lines construct their own narratives may help prevent tragedy in other industries. Additionally, the fallout from the accidents has implications for the field of human-machine communication in general. Machines will inevitably fail, sometimes with tragic consequences. What can these incidents tell us about the human-machine relationship under stress? And, what can aviation's experience teach us about the evolving tension between humans and machines in related industries and transport technologies?

Deskilling Machines

We find it compelling that throughout over 5,000 posts, we never observed the word *skill* being used to describe machines, but the same word is used countless times to describe human capabilities. It's also telling that users often refer to manual flying skills as *hand-flying*. The term almost seems purposefully chosen to exclude a machine from eligibility; only humans have hands. Guzman (2020) identifies a number of ontological boundaries between humans and machines, including emotion, intelligence, and autonomy. It is unclear where the notion of skill may fit in, or perhaps—in the wake of two devastating accidents—pilots recast the meaning of skill to provide a new barrier between human and machine. Additionally, maybe retaining skill as human-only also serves to defend the honor of the profession and pilots' sense of professional competence that they feel is being eroded by manufacturers, regulators, and corporate leadership. Regardless, what is important is to observe how the effects of industry traditions, context, and perhaps a dose of human ego, can lead to the collective drawing of new boundaries between human and machine in unexpected places. This finding raises questions about related transport technologies such as autonomous vehicles (AVs)—will tomorrow's self-driving cars be referred to as skilled in marketing materials or in public discourse? How might the verbiage change following accidents? This is a promising area for future research given the rapid development in AV technology.

Human-Machine Uncommunication

Almost no aspect of the accidents generated more outrage from aviation professionals than the secretive implementation and operation of the MCAS system. The field of human-machine communication assumes that machines act as interlocutors in a number of capacities (i.e., agency, influence, interactivity (Banks & Graaf, 2020)). However, has enough consideration been given to machines that do not communicate? Is a system like MCAS outside the boundaries of the discipline—and if so, how can we participate in the design of better systems in the future? We definitely see a place for human-machine communication research in machines that are not obviously communicative (e.g., the "mute machines" and industrial robots studied by Guzman, 2016) and more research should be directed at uncommunicative machines.

The anger over the silence of the machine is also important to other industry leaders looking to implement automation, especially in other transport technologies like AVs. Perhaps not every action by a machine must be communicated; information overload can be just as harmful as silence. But machine actions that override human inputs should be announced. We see consistently in our results that pilots feel their autonomy is challenged by flightdeck automation. So it is embarrassing enough to humans that a machine may be trusted more to perform certain actions, failure to inform humans of this may cultivate an atmosphere of distrust that destroys relationships between human and machine.

Enduring Tensions

In a recent presentation, Professor Rich Ling highlighted the ongoing tension that exists between human and machine (Ling, 2021). In our study we see it manifest as competition. Remarkably, there is only token mention of human-machine collaboration—on the flightdeck it is a zero-sum game in the battle for agency. We also see the tension manifest between different groups of users: some who see the machines as inevitable and accept their fate, others who maintain that humans are still the ultimate authority. This is a fruitful avenue for future research as machines advance into personal and professional life. Will there be a day where some have ceded control of aspects of their life to machines? After all, some would likely argue we are "already there" when we let Netflix choose our next movie night or Yelp tells us where to eat.

We also witness human-machine tension as pilots negotiate their changing professional identity. The lesson for industry leaders is to pay close attention when implementing machines as they may disrupt employee's sense of pride and purpose, ultimately disrupting communication channels between employees and management and—in aviation—compromising safety. The lesson for human-machine communication scholars is all of the above, but with an emphasis on the need to further study the factors that lead to both the resentment *of* and resignation *to* machines. We end with an encapsulating post:

> These two accidents are a perfect example of a problem that is only going to get worse. You can't make better pilots by putting them in airplanes that fly themselves until the day they don't, then expecting them to fly it out of the fire. The days of good hands and feet flying are never coming back, and this is not news to

those who make the software. The sooner we give up and turn it all over to them the better. You can rest assured they are ready for it, even if the public and the pilots are not. That said, if it was me getting kicked out of the loop they would have to pry my cold dead hands off the controls.

Conclusion

In 2019, two aviation tragedies reignited the long-running debate of human versus machine. In witnessing the professional aviation community make sense of the incidents we see narratives emerge that describe human versus machine as a zero-sum game, as already lost, and as a battle which has poisoned an entire profession with bitterness and distrust. The pilots of Flight 610 and 302 were ultimately unsuccessful against machine in their fight for flight; but their actions have spurred a discussion which offers a window into the future of the complicated relationship between human and machine. A decade from now, we are sure to be witnessing these same tensions in the countless other industries implementing automation. At that time, we may not look back to aviation and think "we've seen this movie before," but we can rest assured that aviation and the 737 MAX incidents provide a sneak preview.

Author Biography

Andrew Prahl (PhD University of Wisconsin-Madison) is an Assistant Professor at the Wee Kim Wee School of Communication & Information at Nanyang Technological University, Singapore. His research addresses the communication consequences of replacing humans with machines. More broadly, Andrew's research investigates the key issues raised by the automation of labor.

https://orcid.org/0000-0003-3675-3007

References

Banks, J., & Graaf, M. de. (2020). Toward an agent-agnostic transmission model: Synthesizing anthropocentric and technocentric paradigms in communication. *Human-Machine Communication*, *1*(1). https://doi.org/10.30658/hmc.1.2

Battiste, V., Lachter, J., Brandt, S., Alvarez, A., Strybel, T. Z., & Vu, K.-P. L. (2018). Human-automation teaming: Lessons learned and future directions. In S. Yamamoto & H. Mori (Eds.), *Human interface and the management of information. Information in applications and services* (pp. 479–493). Springer International Publishing. https://doi.org/10.1007/978-3-319-92046-7_40

Billings, C. E. (1996). *Human-centered aviation automation: Principles and guidelines*. National Aeronautics and Space Administration, Ames Research Center.

Black, L. W. (2008). Deliberation, storytelling, and dialogic moments. *Communication Theory*, *18*(1), 93–116. https://doi.org/10.1111/j.1468-2885.2007.00315.x

Calhoun, G. (2021). Adaptable (Not adaptive) automation: The forefront of human–automation teaming. *Human Factors*, 00187208211037457. https://doi.org/10.1177/00187208211037457

Carlson, M. (2015). The robotic reporter. *Digital Journalism*, *3*(3), 416–431. https://doi.org/10.1080/21670811.2014.976412

Driskill, M. (2021, June 19). Viewpoint: Single-pilot operations: Just say no. *Asian Aviation*. https://web.archive.org/web/20210619230923/https://asianaviation.com/viewpoint-single-pilot-ops-just-say-no/

Erlingsson, C., & Brysiewicz, P. (2017). A hands-on guide to doing content analysis. *African Journal of Emergency Medicine*, *7*(3), 93–99. https://doi.org/10.1016/j.afjem.2017.08.001

Farvre, C. (1994). Fly-by-wire for commercial aircraft: The Airbus experience. *International Journal of Control*, *59*(1), 139–157. https://doi.org/10.1080/00207179408923072

Fortunati, L., & Edwards, A. (2021). Moving ahead with human-machine communication. *Human-Machine Communication*, *2*(1). https://doi.org/10.30658/hmc.2.1

Franzke, A., Bechmann, A., Zimmer, M., & the Association for Internet Researchers. (2020). *Internet research: Ethical guidelines 3.0.* https://aoir.org/reports/ethics3.pdf

Frost, L. (2021, June 17). EXCLUSIVE Cathay working with Airbus on single-pilot system for long-haul. *Reuters*. https://web.archive.org/web/20210617090817/https://www.reuters.com/business/aerospace-defense/exclusive-cathay-working-with-airbus-single-pilot-system-long-haul-2021-06-16

Gates, D. (2019). The inside story of MCAS: How Boeing's 737 MAX system gained power and lost safeguards. *The Seattle Times*. https://web.archive.org/web/20190622230059/https://www.seattletimes.com/seattle-news/times-watchdog/the-inside-story-of-mcas-how-boeings-737-max-system-gained-power-and-lost-safeguards/

Gibbs, J., Kirkwood, G., Fang, C., & Wilkenfeld, J. (2021). Negotiating agency and control: Theorizing human-machine communication from a structurational perspective. *Human-Machine Communication*, *2*(1). https://doi.org/10.30658/hmc.2.8

Guzman, A. (2016). The messages of mute machines: Human-machine communication with industrial technologies. *Communication 1*, *5*(1), 1–30. https://doi.org/10.7275/R57P8WBW

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

Helmore, E. (2019, June 17). Profit over safety? Boeing under fire over 737 Max crashes as families demand answers. *The Guardian*. https://web.archive.org/web/20190617081715/https://www.theguardian.com/business/2019/jun/17/boeing-737-max-ethiopian-airlines-crash

Ibsen, A. Z. (2009). The politics of airplane production: The emergence of two technological frames in the competition between Boeing and Airbus. *Technology in Society*, *31*(4), 342–349. https://doi.org/10.1016/j.techsoc.2009.10.006

KMKT. (2019). *Final aircraft accident investigation report; KNKT.18.10.35.04; PT. Lion Mentari Airlines Boeing 737-8 (MAX); PK-LQP Tanjung Karawang* [Final]. Komite Nasional Keselamatan Transportasi.

Ling, R. (2021, June 14). *Tensions in the adoption of AI: Thinking with Georg Simmei on human vs mechanical, and rationalisation vs social cohesion* [Seminar].

Moreno, M. A., Goniu, N., Moreno, P. S., & Diekema, D. (2013). Ethics of social media research: Common concerns and practical considerations. *Cyberpsychology, Behavior, and Social Networking*, *16*(9), 708–713. https://doi.org/10.1089/cyber.2012.0334

Neuendorf, K. A. (2002). *The content analysis guidebook*. SAGE.

O'Sullivan, M. (2017). The untold story of QF72: What happens when "psycho" automation leaves pilots powerless? *The Sydney Morning Herald*. https://web.archive.org/web/20180317025236/https://www.smh.com.au/lifestyle/the-untold-story-of-qf72-what-happens-when-psycho-automation-leaves-pilots-powerless-20170511-gw26ae.html

Piercy, C., & Gist-Mackey, A. (2021). Automation anxieties: Perceptions about technological automation and the future of pharmacy work. *Human-Machine Communication*, *2*(1). https://doi.org/10.30658/hmc.2.10

Ruskin, K. J., Corvin, C., Rice, S. C., & Winter, S. R. (2020). Autopilots in the operating room: Safe use of automated medical technology. *Anesthesiology*, *133*(3), 653–665. https://doi.org/10.1097/ALN.0000000000003385

Sarter, N. B., & Woods, D. D. (1997). Team play with a powerful and independent agent: Operational experiences and automation surprises on the Airbus a-320. *Human Factors*, *39*(4), 553–569. https://doi.org/10.1518/001872097778667997

Smith, B. (2016). Narrative analysis. In E. Lyons & A. Coyle (Eds.), *Analysing qualitative data in psychology* (2nd ed., pp. 202–221). Sage. https://eprints.kingston.ac.uk/id/eprint/31580

Sullivan, K. (2019). "I've become very isolated": The aftermath of near-doomed QF72. *The Sydney Morning Herald*. https://web.archive.org/web/20190517070057/https://www.smh.com.au/national/i-ve-become-very-isolated-the-aftermath-of-near-doomed-qf72-20190514-p51n7q.html

Tsang, P. S., & Vidulich, M. A. (2002). *Principles and practice of aviation psychology*. CRC Press.

Appendix: Total Codes Count

Technical Discussions			
#	**Code**	**N**	**Percent**
1	Aircraft Systems	540	13.61
2	MCAS	276	6.96
3	Flight Data Recorder (FDR)	115	2.90
4	Operational Procedures	82	2.07
5	Flight Tracking Software	58	1.46
6	Principles of Flight (POF)	50	1.26
7	Data Analysis	44	1.11
8	Simulators	42	1.06
9	Switch Design	22	0.55
10	Fly-By-Wire (FBW)	16	0.40
	Total	**1,245**	**31.38**

Investigative Discussions			
#	**Code**	**N**	**Percent**
1	Aircrew	409	10.31
2	Corrective Measures	298	7.51
3	Boeing	280	7.06
4	Emergency Grounding	192	4.84
5	Investigation Board	139	3.50
6	Regulators	108	2.72
7	Ethiopian Airlines	76	1.92
8	Lion Air	60	1.51
9	Service Resumption	15	0.38
10	Cancelled Orders	3	0.08
11	Lawsuits	1	0.03
12	Aircraft Registration	1	0.03
	Total	**1,582**	**39.88**

Comparative Discussions			
#	**Code**	**N**	**Percent**
1	Other Incidents	71	1.79
2	Other Manufacturing Companies	32	0.81
3	Other Aircraft	24	0.60
4	Other Airline Companies	13	0.33
	Total	**140**	**3.53**

General Discussions			
#	**Code**	**N**	**Percent**
1	Aviation Industry	17	0.43
2	Thought Experiment	11	0.28
3	Summary of Happenings	4	0.10
	Total	**32**	**0.81**

Miscellaneous Comments			
#	**Code**	**N**	**Percent**
1	Aviation Industry	199	5.02
2	Irrelevant Personal Responses	336	8.47
3	Unrelated Questions	183	4.61
4	Irrelevant Argumentations	89	2.24
5	Expressions of Formality	59	1.49
6	Speculatory Statements	48	1.21
7	External Resources	31	0.78
8	Emotional Responses	12	0.30
9	Repeated Posts	4	0.10
10	Indecipherable Posts	7	0.18
	Total	**968**	**24.40**

Human-Machine Communication Scholarship Trends: An Examination of Research From 2011 to 2021 in Communication Journals

Riley J. Richards[1] , Patric R. Spence[2] , and Chad C. Edwards[3]

1 Department of Communication, Oregon Institute of Technology, Klamath Falls, Oregon, USA
2 Nicholson School of Communication and Media, University Central Florida, Orlando, Florida, USA
3 School of Communication, Western Michigan University, Kalamazoo, Michigan, USA

Abstract

Despite a relatively short history, the modern-day study of communication has grown into multiple subfields. To better understand the relationship between Human-Machine Communication (HMC) research and traditional communication science, this study examines the published scholarship in 28 communication-specific journals from 2011–2021 focused on human-machine communication (HMC). Findings suggest limited prior emphasis of HMC research within the 28 reviewed journals; however, more recent trends show a promising future for HMC scholarship. Additionally, HMC appears to be diverse in the specific context areas of research in the communication context. Finally, we offer future directions of research and suggestions for the development of HMC.

Keywords: human-machine communication, HMC, journals, communication studies, human-robot interaction

Communication has a long history of examining best practices in various contexts. Most scholars trace the beginnings of the Communication Studies/Science field to the rhetorical traditions of the Greeks, the work of Shannon and Weaver, sociology, and others examining propaganda in World War II (Berger, 1991; Schramm et al., 1959; Song et al., 2020). Despite

CONTACT   Riley J. Richards · Department of Communication · Oregon Institute of Technology · 3201 Campus Dr., Klamath Falls, OR 97601 · riley.richards@oit.edu

a relatively short history, the modern-day study of communication has grown into multiple subfields. For example, the International Communication Association (ICA) includes 33 divisions and interest groups. Mass communication is ICA's longest-standing division, and Human-Machine Communication is the most recent (Tenenboim-Weinblatt & Lee, 2020). Even within the breadth of the communication field, significant differences can be found in the larger and smaller contexts (e.g., interpersonal, organizational, health) (Erba et al., 2018). Newer fields have grown out of even relatively newer subfields. For example, crisis communication is growing out of organizational communication (Lachlan et al., 2019). The communication discipline and subfields are ever-changing.

The newest subfield, Human-Machine Communication (HMC), has developed out of many subfields both in communication and with sympathies to related fields. Most of the work examining machine actors is represented in other fields. Research in human-computer interaction (HCI), human-robot interaction (HRI), and human-agent interaction (HAI) have flourished in the last decades. The discipline of Communication Science/Studies has focused primarily on computer-mediated communication (CMC: technology mediates the interaction between people) (Westerman et al., 2020). The critical point in this research is that the channel of communication is mediated by technology. To better understand the relationship between human-machine communication research and traditional communication science, this study examines the published scholarship in communication-specific journals from 2011–2021 that are focused on HMC. In the following sections, the meaning of HMC and what research areas have been examined under the HMC umbrella are discussed. Specifically, HMC articles from a population of over 9,000 published articles are selected for further inquiry. The study then concludes through an analysis of the trends that emerge in HMC publications for the last decade.

What Is Human-Machine Communication?

Although communication technologies have long been studied as media or channels through which people communicate with one another, machine actors are increasingly being used as a source of communication for other humans. The subfield of HMC has developed to examine this space where machine actors can be communication partners. Although "communication with digital interlocutors ontologically is not the same thing as communication with another human," there are helpful human communication theories and paradigms to understanding HMC (Fortunati & Edwards, 2020, p. 7). As Westerman et al. (2020) note about the role of machine actors in communication, "social action is social action" (p. 394) and thus can be examined using traditional human communication theories as a starting place. Guzman (2018) argues that HMC is not a competitor to HCI, HAI, or HRI but rather subsumes them when communication is central to the investigation. She notes that "HMC can be thought of as an umbrella encompassing the many approaches to people's communication with various technologies" (Guzman, 2018, p. 7). These technologies "are not a medium through which humans interact, but rather a medium with which humans interact" (Zhao, 2006, p. 402). A. Edwards and Edwards (2017) note that "HMC involves communication with digital interlocutors including embodied machine communicators, virtual and artificially intelligent agents (e.g., spoken dialogue systems), and technologically

augmented persons, either in real or virtual and augmented environments" (p. 487). As such, technology is viewed as more than a channel or medium in HMC; it takes on the role of a communicator. This is the crucial distinction between HMC and CMC. While CMC focuses on technology as a medium between people, HMC focuses on a machine potentially serving as a communication partner.

Spence (2019) asked if there was a theory central to the study of HMC. Questions concerning a central theory in HMC have sparked many discussions, bolstering the case for treating HMC as a distinct subject. The interdisciplinary nature of this work is vital for HMC to help develop as a field (Fortunati & Edwards, 2021). In HMC, several existing ideas from other disciplines function well to answer critical questions (See Spence, 2019 for more details). Despite the examination of existing theories from diverse subfields, there is the development of theories specifically targeted at HMC (Craig & Edwards, 2021; A. Edwards et al., 2019; C. Edwards, Edwards, et al., 2016; Spence et al., 2014) and extensions on the CASA paradigm (Gambino et al., 2020; Lombard & Xu, 2021). More targeted and specific theories will be developed and tested as the field grows related to the HMC. We argue that this is necessary for continued growth.

Examining HMC Scholarship from 2011 to 2021

HMC has thus far followed a similar path to other older subfields, growing out of multiple communication fields as previously discussed. The HMC interest group at ICA was officially recognized in 2019. Early-career scholars are beginning to be formally trained in the field. More senior scholars have transitioned their former training and research agendas into the new HMC field. Since the beginning, HMC has grown each year regarding the number of submissions to the interest group. The *Human-Machine Communication* journal was started in 2020 as an open-access journal with an international editorial board to provide a central place for discussion and promotion of research. There have been two HMC workshops at IEEE's HRI conference, workshops and panels at AoIR (Association of Internet Researchers), special issues devoted to HMC in the Central States Communication Association's *Communication Studies*, the *Journal of Communication Pedagogy* journals, and preconferences about HMC at ICA. In short, the last few years have produced a massive increase in the amount of attention devoted to HMC. The same can be said for the areas of HMC research as well.

HMC has been expanding, incorporating subfields of communication. For example, instructional communication scholars have studied robots as teachers (e.g., Abendschein et al., 2021; C. Edwards, Edwards, et al., 2016; C. Edwards et al., 2021). Scholars have examined social robots as interpersonal relational partners (A. Edwards et al., 2019; Ling & Björling, 2020; Lutz & Tamò-Larrieux, 2020; Mattiassi et al., 2021; Rodríguez-Hidalgo, 2020) and chatbots (Banks & Van Ouytsel, 2020; Beattie et al., 2020; C. Edwards, Beattie, et al., 2016). Journalism scholars consider the implications for A.I. reporters (Carlson, 2015; Johanssen & Wang, 2021; Lewis et al., 2019), and crisis and strategic communication scholars consider behavioral outcomes from news from a human or robot (Rainear et al., 2021; Rainear et al., 2019). Moreover, scholars have been interested in issues of identity as it relates to HMC (Davis & Stanovsek, 2021; Dehnert & Leach, 2021; Guzman, 2020; Liu, 2021). Some scholars have focused on issues of automation and algorithms (Ishii et al., 2021; Piercy &

Gist-Mackey, 2021; Utz et al., 2021). Others have examined HMC regarding ethics and theories of mind (Banks, 2019, 2020; Gunkel, 2018, 2020). The larger context (e.g., instruction, interpersonal) and specific HMC context (e.g., humanoid, chatbot) are essential to consider because part of advancing any field is to recognize the collective efforts of its smaller parts (Berger, 1991). These contexts are examined for the current project.

HMC scholarship published from 2011 to 2021 was the central focus of the current study. This time frame was chosen to allow for a glimpse of the development of this work from near its beginning. Although there are earlier examples of HMC work (e.g., Rau et al., 2009; Zhao, 2006), this window demonstrates when mainstream communication science/studies journals started to publish HMC work. Additionally, this 11-year period coincides with the development of technology utilized in HMC research. For example, Softbank's popular NAO robot was used extensively during this time frame. The same company produced a more robust social robot, Pepper, in 2015. For voice-based agents (VBAs), Amazon's Alexa was released in 2014, which spurred many studies examining this type of machine actor. The first consumer model of Google Glasses and the Oculus VR system started selling in 2014 and 2016, respectively. In other words, our time frame for the current study captures the availability of many of the research tools used. Furthermore, this 11-year window has been used by similar work examining the development of a new field (Beck et al., 2004; Lachlan et al., 2019; Spence & Baker, 2007).

HMC is an interdisciplinary field and focuses on the communication aspects of machine actors. Although relatively new, it has roots in early communication research but expands to new contexts. Examining the published scholarship over the last decade will allow researchers to see opportunities for gaps in the literature, understand how existing HMC research is being conducted, and what types of methods are being used. We will present a series of research questions addressing these issues in the following sections. The first set of research questions will examine where HMC research is published and what context.

> **RQ1:** What percentage of regional, national, and international communication journal articles host human-machine communication research?

> **RQ2:** Among the human-machine communication journal articles, what is the (a) larger and (b) specific human-machine communication context?

Just as scholars have carried over prior studied contexts of interest, they have transitioned traditional human-to-human communication research methodologies over to HMC. So much so, the inaugural article of the HMC journal was titled "Opening space for theoretical, methodological, and empirical issues in human-machine communication" (Fortunati & Edwards, 2020, p. 1). Methodology of interest generally includes quantitative, qualitative, rhetorical, or some combination of the broad methodology (i.e., mixed-methods). Even within a seemingly homogeneous, more extensive method such as quantitative, a specific sub-method like survey can vary greatly. For example, the channel the survey is distributed through (e.g., asynchronous online, synchronously administered by the researcher over the phone or in-person), when it is distributed (e.g., cross-sectional, longitudinal over years or days), and who it is distributed to (e.g., one person, dyad, students, workforce). These small but meaningful differences in study design can significantly impact

individual research studies, thus making meta-analyses which are an important method, difficult to apply for now (Schmidt & Hunter, 2015).

Traditional meta-analyses can be valuable tools in correcting study artifacts (e.g., sampling error, error of measurement) and providing a comprehensive review of quantitative results. However, traditional meta-analyses cannot summarize nonempirical arguments. HMC research is influenced by various epistemological and methodological concerns (Guzman & Lewis, 2020), including nonempirical scholarship. For example, HMC scholars have applied rhetorical (Coleman, 2021) and philosophical approaches to machine ethics (e.g., Gunkel, 2012). Therefore, we choose to be inclusive rather than exclusive of our colleagues' work when considering the following research question. We do not know beyond anecdotal evidence of the diverse research methodology techniques, including settings and participant samples used within HMC research. Such knowledge will highlight potential research gaps and point to future research directions. Therefore, the third research question is given.

> **RQ3**: Among the human-machine communication journal articles, what (a) methodological techniques, (b) settings, and (c) samples are most often used?

Method

A census of articles in 28 communication journals spanning 11 years (2011–2021) was completed, resulting in examining 132 HMC articles. To be included, an article was required to be a regular article, editorial, or part of a colloquium. Statements from association presidents, errata, and book reviews were not included in the article count. From the 9,497 articles that fit that criterion, articles then had to be classified as either HMC in focus or not. To be classified as an HMC article, the coder needed to identify the article as fitting into the following criteria: (a) "Human-machine communication as a process is an exchange of messages between people and technology, but in the course of the interaction and as a result of it, both the machine and human may also take on other roles" (Guzman, 2018, p. 17), (b) "HMC involves communication with digital interlocutors including embodied machine communicators, virtual and artificially intelligent agents (e.g., spoken dialogue systems), and technologically augmented persons, either in real or virtual and augmented environments" (A. Edwards & Edwards, 2017, p. 487), and (c) the author(s) identified the article as dealing with HMC in the title, abstract, or keywords. From these criteria 132 Human-Machine Communication emerged from the population defined. Articles were then subject to additional coding to address the research questions in this study.

The 28 journals that were selected for analysis were: *Asian Journal of Communication, Communication, Culture and Critique, Communication and Critical/Cultural Studies, Communication Education, Communication Studies, Communication Monographs, Communication Quarterly, Communication Reports, Communication Research, Communication Research Reports, Communication Teacher, Communication Theory, Critical Studies in Media Communication, First Amendment Studies/Communication and Democracy, Human Communication Research, Human-Machine Communication, Journal of Applied Communication Research, Journal of Communication Pedagogy, Journal of Communication, Journal of Computer Mediated Communication, Journal of International and Intercultural Communication,*

Mobile Media & Communication, *New Media & Society*, *Quarterly Journal of Speech*, *Review of Communication*, *Southern Communication Journal*, *Text and Performance Quarterly*, and *Western Journal of Communication*. The selection of periodicals included in this analysis was based on previous studies and literature in citation analysis (see Bolkan et al., 2012; Griffin et al., 2018; Hickson et al., 2009; Lachan et al., 2019; Spence & Baker, 2007) and were chosen to give the most substantial representation to HMC within the literature in the field of communication. Whereas other journals may have HMC articles represented, their absence of a direct focus on communication removed them from the analysis.

Measures

Type of Measurement

Articles were classified broadly based on the measurement used; therefore, classifications included quantitative measures, qualitative measures, mixed measures/methods, or no measures. If an article was classified as HMC, subsequent coding determined whether it was quantitative when using meaningful numeric symbols. In contrast, an article was considered qualitative if it presented and discussed results coming from methods such as focus groups, thematic analysis, interviews, or ethnography. An article was coded as mixed measurement if a combination of quantitative or qualitative measures was used as indicated above. A study was coded as having no measurement if there was no attempt to employ an analysis from any criteria or system of measurements; articles fitting this criterion include theory development or literature reviews.

Data Collection and Setting

The procedures of data collection and the research setting also were explored. Eleven categories for how data were collected emerged from discussions between the coders and influence of previous studies (Lachlan et al., 2019; Spence & Baker, 2007). These included experiments, survey research, archival data (including rhetorical analysis, case studies, content analysis, and other texts), meta-analysis, interviews, ethnography, focus groups, literature reviews (including commentaries and theory development), observation, and multiple data collection procedures.

For the research setting, 10 categories were used; these included laboratory research, online experimental research, mail survey, human-assisted surveys, online survey research, online data collection (including social media harvesting), focus groups, field research (including interviews outside a laboratory), multiple settings, and no research setting (such as literature reviews or textual analysis).

Context of the Study

The specific context of both the study and the context of HMC were examined. Ten contexts emerged from discussions between two coders; these included studies focused on interpersonal relationships, small groups, organizational studies, education studies including classroom and instructional communication, journalism, health care, intercultural studies, mass media, rhetoric or analysis of a text, and general communication studies.

Specific context to the role of HMC included the following nine categories: Human-Robot Interaction (HRI), Human-Agent Interaction (HAI), Virtual or Augmented Reality, Automation, Human-Computer Interaction (HCI), Artificial Intelligence (A.I.), General Studies, Algorithms, and Cybernetics.

Participants Within the Examined Studies

Also under investigation were the participants that made up the examined studies. These included student samples, completely volunteer/compensation not specified, social media recruitment, mechanical recruitments (such as MTurk/Qualtrics), multiple recruitment methods, no participants, or participants that receive financial compensation (without mechanical recruitment). The decision to have mechanical recruitment, which involves paid compensation, as a category separate from monetary compensation was based on two reasons provided by Lachlan et al. (2019): These recruitment methods are increasing in popularity, and the debate surrounding their use (see Kees et al., 2017; Sheehan & Pittman, 2016) provides a good avenue of investigation. Moreover, this analysis contributes to a base-line of using this type of recruitment which is distinct from other forms of paid research participation.

Additional Measures

Other information collected was author name and affiliation, Altmetric score, and citation count from Crossref.

Intercoder Reliability

To determine the reliability of the coding of the articles and content between the three coders, a reliability check was completed. Three coders completed categorization of all content within *Communication Research Reports* for the 11-year period. This coding yielded 450 articles for the reliability check concerning categorizing the articles as HMC. A perfect agreement was found between the three coders concerning if an article was classified as an HMC article or not. Then two of the coders completed an analysis of all categories of interest in the 14 articles that were coded as HMC. Intercoder reliability was assessed using ReCal2 (Freelon, 2010). Perfect agreement (Scott's Pi = 1.0) was found for all variables involving coder judgments: general context, HMC context, research setting, type of measurement, data collection, and participants. After perfect agreement was met, the authors independently coded the remaining journals: first author ($N = 13$), second author ($N = 2$), and third author ($N = 12$).

Primary Results

To answer the first research question inquiring the percentage of regional, national, and international communication journal articles devoted to HMC research, several analyses were completed. Full descriptive statistics are reported in Table 1, Table 2, Table 3, and Table 4.

The number of HMC articles was first determined in the census of 9,497 entries published in the investigated time period. Approximately 1.39% of the articles were classified as HMC ($N = 132$). *Human-Machine Communication* published the most articles ($N = 25$) and had the highest percentage of HMC articles published (100%). In reference to the number of HMC articles published, the top four journals after *Human-Machine Communication* were *New Media & Society* ($N = 21$), *Communication Studies* ($N = 20$) with *Communication Research Reports* ($N = 13$), and *Journal of Computer Mediated Communication* following ($N = 13$). By percentage of articles, and after the journal *Human-Machine Communication*, it was again *Communication Studies* and *Journal of Communication Pedagogy* (4.5%) with the *Journal of Computer Mediated Communication* (4.1%) and *Communication Research Reports* (2.9%) following. Both *Communication Studies* and the *Journal of Communication Pedagogy* had special issues that contributed to their numbers during the examination period. Of the 132 articles that were classified as HMC, 26, or 19.7%, appeared in some type of special issue (not necessarily a special issue related to HMC).

HMC articles began to receive increased recognition in the literature starting in 2018 when the number of published articles tripled and began an ascent in representation. In addition, the year 2021 had the highest number of HMC articles published ($N = 57$), followed by 2020 ($N = 26$), 2018 ($N = 14$), and 2019 ($N = 12$). This increase in published articles began 2 years before the launch of the *Human-Machine Communication* journal and highlights both the need for the journal and the emergence of the field of study.

The number of citations an article received and the Altmetric score of each article was also examined. Article citations were based on Crossref citations listed on the journal landing page for each article if available. At the time of data collection, the three most cited articles appeared in the *Journal of Communication*, *New Media & Society*, and *Communication Research Reports* (Ho et al., 2018; Guzman & Lewis, 2020; Spence, et al., 2014, with 79, 78, and 69 citations, respectively). For Altmetric scores, two of the three top scores also appeared in *New Media & Society* and one in *Critical Studies in Media Communication* (Woods, 2018; Guzman & Lewis, 2020; and Yan et al., 2021 with scores of 143, 111, and 78, respectively). The article by Guzman & Lewis (2020) in *New Media & Society* is represented in both the top 3 of citations and Altmetric scores of HMC articles, highlighting this article's heuristic provocation and possibly the perception of interest from readers.

The second research question focused on the context of both the larger scope of the selected studies and the HMC context itself. When examining the context of the study itself, results reveal that interpersonal communication/relationships are the most examined area of HMC, accounting for 32.6% ($N = 43$) of the articles published. An intercultural context accounted for 22% ($N = 29$) of the articles. Mass media (14.4%, $N = 19$), health care (9.1%, $N = 12$), and education (8.3%, $N = 11$), including instructional communication and classroom settings, were the contexts with the most frequency of analysis.

The specific HMC context analysis revealed that virtual/augmented reality (25.7%, $N = 34$) was the most examined context. Followed closely by human-robot interaction (18.9%, $N = 25$), and human-agent interaction (15.9%, $N = 21$) with artificial intelligence (13.6%, $N = 18$) and general studies (13.6%, $N = 18$) also receiving increased attention.

The third research question concerned the methodological techniques, data collection settings, and participants' characteristics in HMC studies. Concerning measurement of data, studies using quantitative data collection methods accounted for almost half of all

studies (48.5%, *N* = 64), followed by studies using no measurement (35.6%, *N* = 47). Qualitative studies (12.1%, *N* = 16) and mixed-method approaches accounted for the smallest number of studies (3.8%, *N* = 5).

The most used type of data collection was an experiment (40.2%, *N* = 53). The second most type coded in the current study did not collect any kind of data (32.6%, *N* = 43). Studies using archival data (8.3%, *N* = 11) and survey data (8.3%, *N* = 11) had the third-highest frequency of occurrence in the sample. All other categories of data collection were under 5% each.

The place where the research took place was labeled the research setting. Studies that did not involve a research setting, such as articles developing theory or literature review, accounted for 39.4% of all articles (*N* = 52). This was followed by studies taking place in a laboratory (30.3%, *N* = 40) and online experiments (12.1%, *N* = 16), and field research (6.8%, *N* = 9) were represented. All other categories were 3% or less.

Post Hoc Analysis

Due to the recent and rapidly growing HMC field, a baseline of research productivity has yet to be determined. To inform the HMC research community of trends and provide evidence for tenure, promotion, and/or funding agencies, a post hoc analysis of authorship trends was conducted. The study identified prominent authors across the field of HMC. There were 264 different authors within the analysis, with nine authors having three or more publications each and 22 authors having two publications. The analysis included authorship regardless of author order, and therefore in some instances, articles were co-authored by two or more individuals with more than three publications. The results of this analysis suggest clusters of HMC scholars at specific departments. The top nine scholars in the analysis were identified by the total number of publications. Among these authors, three have current or one-time affiliations with Western Michigan University, and three also had been an editor for an HMC-related publication. One scholar with more than three HMC publications was a doctoral student at the time of this analysis. For this analysis, please see the following link: https://osf.io/gh7z2.

Discussion

Machine communicators have rapidly developed in the past decade from humanoids (e.g., NAO and Peppers by Softbank) to VBAs (e.g., Amazon's Alexa) and VR (e.g., Oculus). HMC scholars' research agendas have differed while still falling under the same umbrella (Guzman, 2018). This study reviewed the HMC publications in 28 communication journals over the past decade to consider where HMC as a field has been to understand where the field should go (Berger, 1991).

The results from the first research question show a limited past emphasis (1.39%) within the 28 reviewed journals. However, more recent trends show a promising future for HMC scholarship. The number of HMC-related article publications in the reviewed journals grew yearly. Specifically, from 2019 to 2020, the amount of HMC publications grew by 116% and from 2020 to 2021, it grew 119%. The significant growth is likely due to the HMC interest group being established at ICA in 2019 along with the *Human-Machine*

Communication journal in 2020. Roughly 19% of the reviewed HMC articles were published in the *Human-Machine Communication* journal in 2020–2021. There were meaningful contributions in special issues published in *Communication Studies* (2020 and 2021) and the *Journal of Communication Pedagogy* (2021). These special issues in non-HMC specific journals are expected to educate and potentially attract non-HMC scholars who would not usually read the *Human-Machine Communication* journal. With publication trends and the advancement in machine communication technology, we expect HMC publications to grow as they have in the past 2 years exponentially. To help facilitate this growth, we offer insight into the existing strengths and potential opportunities for growth in terms of context and methodology.

Communication has always been interdisciplinary and communication scholars have taken pride in the fact (Zhu & Fu, 2019). This strength is highlighted by the field's interdisciplinary stances across broader and specific HMC contexts and the epistemological and methodological approaches scholars have applied (e.g., Guzman & Lewis, 2020). However, previous reviews have critiqued the focus of interdisciplinary scholarship in that subfields may not successfully work together (e.g., R. T. Craig, 1999). Additionally, scholars may overly compensate by applying other disciplines' principles (e.g., theories) to communication, and the field will not have its unique perspective (Berger, 1991).

Based on the current study's second research question findings, HMC has defied R. T. Craig's (1999) prediction of drastically diverse fields not being able to work together. Over 60% of the reviewed HMC articles were taken from the relational perspective but not limited to a specific context. The larger contexts of interpersonal (32.6%), intercultural (22%), and health care (9.1%) made up the majority of reviewed studies in comparison to "one-to-many" such as mass media (14.4%) and instructional (8.3%) made up a smaller percentage. Within the specific HMC context, 25.7% of the reviewed articles focused on VR or AR, 18.9% HRI, 15.9% HAI, and 13.6% A.I. No one larger context or HMC-specific context makes up over half of the current field. The current balance between communication context and the machine communicator under consideration is well done. We urge future reviews, like ours, to consider the balance across diverse contexts. If an imbalance is dedicated, this review may offer a spark for minority perspectives to contribute more holistically or encourage a special issue.

In comparison to contexts, the results from the methodological analysis (RQ3a–c) showed an imbalance. Over half (64.4%) of the reviewed articles were empirically-based compared to the nonempirical (35.6%). Such nonempirical articles provided a literature review or begin to generate HMC theory (e.g., A. Edwards et al., 2019; C. Edwards, Edwards, et al., 2016; Gambino et al., 2020; Lombard & Xu, 2021). The present literature reviews and initial HMC theoretical ideas are important and practical to an extent. However, we are reminded of Lewin's (1951) famous quote "there is nothing so practical as a good theory" (p. 169). A fully fleshed-out HMC theory with its theoretical propositions will serve the community well. Such action will further define our field, respond to Berger's (1991) critique of the communication discipline, begin to answer Spence's (2019) question about what the central question HMC scholars are working toward, and open new lines of research inquiry.

Of the reviewed empirical articles, three quarters (75.3%) were quantitative, while qualitative (18.8%) and mixed methods (5.9%) made up a significantly smaller portion. The imbalance is strong but is in line with other related analyses. For example, roughly 85% of

interpersonal scholarships is quantitative, while 13% is qualitative (Braithwaite et al., 2015). Reviews of mobile communication scholarship (a subset of mediated interpersonal scholarship) find 59% of the scholarship is quantitative, while 37% is qualitative and 4% is mixed methods (Kim et al., 2017). When considered together, the strong presence of quantitative methods makes sense given interpersonal was the most common context. Beyond specific communication context, a review of ICA's flagship journal, the *Journal of Communication* revealed roughly 79% of all manuscripts were quantitative. In comparison, 16% were qualitative, and 4% mixed-method (Walter et al., 2018), suggesting HMC methodology aligns with specific contexts and the broader communication field. Or rather, HMC scholars are likely transitioning their former research methods training and applying the same analytical skills to the new HMC context.

More imbalances were found concerning specific methods within the more significant epistemology, thus offering some immediate practical suggestions. The majority (82.8%) of quantitative studies were experiments and surveys, making up a smaller (17.2%) portion. Results are similar but appear to be more experimental focused when compared to other communication technology research (Erba et al., 2018). Of the experiments, three quarters (75.5%) were conducted in-person in a research lab, while the remaining quarter (24.5%) was conducted online. Thus, the majority of our current knowledge stems from cause-and-effect studies. We did not code if individual studies were cross-sectional or longitudinal. Anecdotally, the reviewed experiments considered direct effects instead of lagged or longitudinal. This begs the question, what important nuances of HMC may have been missed that a grounded qualitative (e.g., interviews, observation) or longitudinal quantitative approach would reveal? For example, a user's motive(s) to communicate with a machine (e.g., Choi & Drumwright, 2021) and how, if at all, that changes over time.

HMC research has increased exponentially; however, the present rate offers scholars time to learn a new method or analysis technique and stay up-to-date with the research trends. Intentionally learning a new research method to fulfill a unique project will inherently take more time than using the research methods scholars have formerly used or were formally trained in. Based on our post hoc analysis of authorship trends, the most published HMC scholars, at least in the journals reviewed, such as those in the top 3.8% ($N = 9$), had three or more publications. In comparison, the top 13.1% ($N = 31$) had two or more HMC publications. We expect this metric to provide significant value to our colleagues pursuing academic employment, promotion, tenure, and research funding. Additionally, we hope such data provides evidence of clarity and excitement for junior scholars considering this new and budding field. As for the over 90% of authors who had one HMC publication, we encourage them to keep contributing. Prior, non-HMC reviews have revealed it is more common for authors to publish once within a specific time frame (e.g., Bolkan et al., 2012; Hickson et al., 2009). Although not limited to HMC expressly, understanding why most scholars only publish once on a given topic would be insightful for many reasons.

The present study has clarified the field but is not without its limitations. Given the small sample ($N = 132$) of included articles in the present study, if enough HMC-related articles exist in non-communication-specific journals, it would undoubtedly impact the results. Future similar studies should consist of, or at least consider, HMC articles in non-communication-specific journals. Such change may remove any bias in results such as included scholarship that is HMC but not from communication scholars and thus is

published in their home discipline. Alternatively, our emphasis was on assessing the degree to which regional (Central States, Eastern States, Southern States, and Western States Communication Association), national (National Communication Association), international (ICA) journals, and high-impact communication journals included HMC-related articles. Our focus emphasized the communication aspect of HMC and led us to use specific conceptual definitions of HMC (e.g., A. Edwards & Edwards, 2017; Guzman, 2018). Our results could differ if we expanded upon our inclusion criteria; likewise, they would differ if we based on inclusion criteria by only one definition. Future studies must carefully consider how and why they collect and assess articles and the impact it will have on results. For example, *Computers in Human Behavior* would be one such journal where there is HMC research published.

If we can make two calls to action to our colleagues, it is the importance of theory and innovation. First, acknowledging if certain theories are more common than others and thus steering our understanding and the academic conversation. Unfortunately, the present study did not code for if articles were founded in theory or not (i.e., atheoretical). Such information in future studies would be fruitful. Prior reviews of communication journals have established that 50 to 70% of all articles are atheoretical (Borah, 2017; Walter et al., 2018). Anecdotally, we do not expect HMC articles to have such a high atheoretical ratio; however, that assumption should still be tested. Future studies should examine the theoretical frames chosen by HMC scholars as the community seeks to develop HMC-specific theories.

Second, the lifeblood of any business (Taneja et al., 2016), or in our case academic field, is innovation. We fear if the dominant methodology (laboratory cross-sectional experiment) continues on its trajectory and new scholars solely use and/or are trained in the method, it will lead to naivete in our understanding of HMC. We believe by purposely keeping to our interdisciplinary nature will prevent such an event. Areas of interest that were not discovered in our analysis include, but are not limited to, intersectionality and marginalized individuals and communities (e.g., ethnicity, class, gender identity, sexuality, sexual orientation, physical ability), critical cultural (e.g., prejudice, discrimination), and relational and group development. We suspect these areas to be formally trained in the non-HMC dominant methodology, such as ethnography, focus groups, naturalistic observations, field and case studies, and rhetorical methods. Like all group work, more diverse ideas being contributed through synergistic methods will lead to creative and beneficial outcomes (Towe, 1996).

The purpose of this study was to examine the published scholarship in 28 communication-specific journals from 2011–2021 to uncover the work that has been done in HMC. Findings suggest a strong beginning to this subfield and one that is inclusive of many other subfields in communication. Additionally, this study highlights the need for a HMC-specific journal to foster the continued growth of this research. Future scholars should keep broadening the scope of HMC research into other methodologies (e.g., more qualitative and rhetorical studies) and should seek to develop HMC-specific theories.

Author Biographies

Riley J. Richards (PhD, University of Wisconsin-Milwaukee) is an Assistant Professor of Communication in the Department of Communication at the Oregon Institute of Technology. His research interests include relational communication and behavior from a functionalist perspective, particularly in the context of relational goals, sexual communication, communication technology, taboo topics, and quasi-sexual relationships (e.g., human-robot). Richards's recent research has appeared in the *Journal of Social and Personal Relationships* and *Computers in Human Behavior*.

  ⓘ   https://orcid.org/0000-0003-2612-4063

Patric R. Spence (PhD, Wayne State University) is a Professor at the University of Central Florida. His primary areas of research are crisis communication and social robotics. He is affiliated with the Communication and Social Robotics Labs (www.combotlabs.org).

  ⓘ   https://orcid.org/0000-0002-1793-6871

Chad Edwards (PhD, University of Kansas) is a Professor of Communication in the School of Communication at Western Michigan University. He is a Theodore von Kármán Fellow at RWTH Aachen University. Edwards's research interests include human-machine communication and instructional communication. Recent publications include articles in: *International Journal of Social Robotics, Frontiers in Robotics and AI, Communication Education, Computers in Human Behavior, Journal of Computer-Mediated Communication,* and *Communication Studies*. Chad Edwards co-directs the Communication and Social Robotics Labs (www.combotlab.org). Currently, Chad is the Vice-Chair of the Human-Machine Communication area at the International Communication Association.

  ⓘ   https://orcid.org/0000-0002-1053-6349

References

Abendschein, B., Edwards, C., Edwards, A., Rijhwani, V., & Stahl, J. (2021). Human-robot teaming configurations: A study of interpersonal communication perceptions and affective learning in higher education. *Journal of Communication Pedagogy*, *4*, 123–132. https://doi.org/10.31446/JCP.2021.1.12

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

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

Banks, J., & Van Ouytsel, J. (2020). Cybersex with human-and machine-cued partners: Gratifications, shortcomings, and tensions. *Technology, Mind, and Behavior*, *1*(1). https://doi.org/10.1037/tmb0000008

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, *71*(3), 409–427. https://doi.org/10.1080/10510974.2020.1725082

Beck, C. S., Benitez, J. L., Edwards, A., Olson, A., Pai, A., & Torress, M. B. (2004). Enacting "health communication": The field of health communication as constructed through publication in scholarly journals. *Health Communication*, *16*(4), 475–492. https://doi.org/10.1207/s15327027hc1604_5

Berger, C. R. (1991). Communication theories and other curios (Chautauqua: Why are there so few communication theories?). *Communication Monographs*, *58*(1), 101–113. https://doi.org/10.1080/03637759109376216

Bolkan, S., Griffin, D. J., Holmgren, J. L., & Hickson, M., III. (2012). Prolific scholarship in communication studies: Five years in review. *Communication Education*, *61*(4) 380–394. https://doi.org/10.1080/03634523.2012.699080

Borah, P. (2017). Emerging communication technology research: Theoretical and methodological variables in the last 16 years and future directions. *New Media & Society*, *19*(4), 616–636. https://doi.org/10.1177/1461444815621512

Braithwaite, D. O., Schrodt, P., & Carr, K. (2015). Introduction: Meta-theory and theory in interpersonal communication research. In D. O. Braithwaite & P. Schrodt (2nd ed.), Engaging in theories in interpersonal communication: Multiple perspectives (pp. 1–20). SAGE.

Carlson, M. (2015). The robotic reporter: Automated journalism and the redefinition of labor, compositional forms, and journalistic authority. *Digital Journalism*, *3*(3): 416–431. https://doi.org/10.1080/21670811.2014.976412

Choi, T. R., & Drumwright, M. E. (2021). "OK, Google, why do I use you?" Motivations, post-consumption evaluations, and perceptions of voice AI assistants. *Telematics and Informatics*, *62*, 101628. https://doi.org/10.1016/j.tele.2021.101628

Coleman, M. C. (2021). Leveraging the rhetorical energies of machines: Covid-19, misinformation, and persuasive labor. *Human-Machine Communication*, *3*, 11–26. https://doi.org/10.30658/hmc.3.2

Craig, M. J., & Edwards, C. (2021). Feeling for our robot overlords: Perceptions of emotionally expressive social robots in initial interactions. *Communication Studies*, *72*(2), 251–265. https://doi.org/10.1080/10510974.2021.1880457

Craig, R. T. (1999). Communication theory as a field. *Communication Theory*, *9*(2), 119–161. https://doi.org/10.1111/j.1468-2885.1999.tb00355.x

Davis, D. Z., & Stanovsek, S. (2021). The machine as an extension of the body: When identity, immersion and interactive design serve as both resource and limitation for the disabled. *Human-Machine Communication*, *2*, 121–135. https://doi.org/10.30658/hmc.2.6

Dehnert, M., & Leach, R. B. (2021). Becoming human? Ableism and control in Detroit: Become human and the implications for human-machine communication. *Human-Machine Communication*, *2*, 137–152. https://doi.org/10.30658/hmc.2.7

Edwards, A., & Edwards, C. (2017). The machines are coming: Future directions in instructional communication research. *Communication Education*, *66*(4), 487–488. https://doi.org/10.1080/03634523.2017.1349915

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, C., Beattie, A. J., Edwards, A., & Spence, P. R. (2016). Differences in perceptions of communication quality between a Twitterbot and human agent for information seeking and learning. *Computers in Human Behavior*, *65*, 666–671. https://doi.org/10.1016/j.chb.2016.07.003

Edwards, C., Edwards, A., Albrehi, F., & Spence, P. (2021). Interpersonal impressions of a social robot versus human in the context of performance evaluations. *Communication Education*, *70*(2), 165–182. https://doi.org/10.1080/03634523.2020.1802495

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

Erba, J., Ternes, B., Bobkowski, P., Logan, T., & Liu, Y. (2018). Sampling methods and sample populations in quantitative mass communication research studies: A 15-year census of six journals. *Communication Research Reports*, *35*(1), 42–47. https://doi.org/10.1080/08824096.2017.1362632

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, *1*, 7–18. https://doi.org/10.30658/hmc.1.1

Fortunati, L., & Edwards, A. (2021). Moving ahead with human-machine communication. *Human-Machine Communication*, *2*, 7–28. https://doi.org/10.30658/hmc.2.1

Freelon, D. G. (2010). ReCal: Intercoder reliability calculation as a web service. *International Journal of Internet Science*, *5*(1), 20–33.

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

Griffin, D. J., Bolkan, S., & Dahlbach, B. J. (2018). Scholarly productivity in communication studies: Five-year review 2012–2016. *Communication Education*, *67*(1), 88–101. https://doi.org/10.1080/03634523.2017.1385820

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Gunkel, D. J. (2020). Mind the gap: Responsible robotics and the problem of responsibility. *Ethics and Information Technology*, *22*(4), 307–320. https://doi.org/10.1007/s10676-017-9428-2

Guzman, A. L. (2018). What is human-machine communication, anyway? In A. L. Guzman (Eds.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

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

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Hickson, M. III., Self, W. R., Johnston, J. R., Peacock, C., & Bodon, J. (2009). Prolific research in communication studies: Retrospective and prospective views. *Communication Research Reports*, *26*, 337–346. https://doi.org/10.1080/08824090903293726

Ho, A., Hancock, J. T., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication*, *68*(4), 712–733. https://doi.org/10.1093/joc/jqy026

Ishii, K., Spence, P. R., & Hodges, W. R. (2021). Social presence in computer-based receptionists: Experimental study towards organizational automation. *Communication Reports*, *34*(2), 92–105. https://doi.org/10.1080/08934215.2021.1918199

Johanssen, J., & Wang, X. (2021). Artificial intuition in tech journalism on AI: Imagining the human subject. *Human-Machine Communication*, *2*, 173–190. https://doi.org/10.30658/hmc.2.9

Kees, J., Berry, C., Burton, S., & Sheehan, K. (2017). An analysis of data quality: Professional panels, student subject pools, and Amazon's Mechanical Turk. *Journal of Advertising*, *46*(1), 141–155. https://doi.org/10.1080/00913367.2016.1269304

Kim, Y., Kim, B., Kim, Y., & Wang, Y. (2017). Mobile communication research in communication journals from 1999 to 2014. *New Media & Society*, *19*(10), 1668–1691. https://doi.org/10.1177/1461444817718162

Lachlan, K. A., Spence, P. R., Seeger, M., Gilbert, C., & Lin, X. (2019). Crisis communication in context: History and publication trends. *Journal of the Association for Communication Administration*, *38*(2), 39–56.

Lewin, K. (1951). *Field theory in social science: Selected theoretical papers*. Harper & Row.

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*(4), 409–427. https://doi.org/10.1080/21670811.2019.1577147

Ling, H., & Björling, E. (2020). Sharing stress with a robot: What would a robot say? *Human-Machine Communication*, *1*, 133–159. https://doi.org/10.30658/hmc.1.8

Liu, J. (2021). Social robots as the bride? Understanding the construction of gender in a Japanese social robot product. *Human-Machine Communication*, *2*, 105–120. https://doi.org/10.30658/hmc.2.5

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.30658/hmc.2.2

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

Mattiassi, A. D. A., Sarrica, M., Cavallo, F., & Fortunati, L. (2021). What do humans feel with mistreated humans, animals, robots, and objects? Exploring the role of cognitive empathy. *Motivation and Emotion*, *45*, 543–555. https://doi.org/10.1007/s11031-021-09886-2

Piercy, C. W., & Gist-Mackey, A. N. (2021). Automation anxieties: Perceptions about technological automation and the future of pharmacy work. *Human-Machine Communication*, *2*, 191–208. https://doi.org/10.30658/hmc.2.10

Rainear, A. M., Jin, X., Edwards, A., Edwards, C., & Spence, P. R. (2021). A robot, meteorologist, and amateur forecaster walk into a bar: Examining qualitative responses to a weather forecast delivered via social robot. *Communication Studies*, 72(6), 1129–1145. https://doi.org/10.1080/10510974.2021.2011361

Rainear, A. M., Lachlan, K. A., & Fishlock, J. (2019). Exploring retention and behavioral intentions when using social robotics to communicate a weather risk. *Computers in Human Behavior*, *90*, 372–379. https://doi.org/10.1016/j.chb.2018.08.029

Rau, P. P., Li, Y., & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587–595. https://doi.org/10.1016/j.chb.2008.12.025

Rodríguez-Hidalgo, C. (2020). Me and my robot smiled at one another: The process of socially enacted communicative affordance in human-machine communication. *Human-Machine Communication*, *1*(1), 55–69. https://doi.org/10.30658/hmc.1.4

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). SAGE.

Schramm, W., Riesman, D., & Bauer, R. A. (1959). The state of communication research: Comment. *The Public Opinion Quarterly*, *23*(1), 6–17.

Sheehan, K. B., & Pittman, M. (2016). *Amazon's Mechanical Turk for academics: The HIT handbook for social science research*. Melvin & Leigh, Publishers.

Song, H., Eberl, J. M., & Eisele, O. (2020). Less fragmented than we thought? Toward clarification of a subdisciplinary linkage in communication science, 2010–2019. *Journal of Communication*, *70*(3), 310–334. https://doi.org/10.1093/joc/jqaa009

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., & Baker, C. R. (2007). State of the method: An examination of level of analysis, methodology, representation and setting in current organizational communication research. *Journal of Northwest Communication Association*, *36*, 111–124.

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*(3), 272–280. https://doi.org/10.1080/08824096.2014.924337

Taneja, S., Pryor, M. G., & Hayek, M. (2016). Leaping innovation barriers to small business longevity. *Journal of Business Strategy*, *37*(3), 44–51. https://doi.org/10.1108/JBS-12-2014-0145

Tenenboim-Weinblatt, K., & Lee, C. J. (2020). Speaking across communication subfields. *Journal of Communication*, *70*(3), 303–309. https://doi.org/10.1093/joc/jqaa012

Towe, L. (1996). *Why didn't I think of that? Creativity in the workplace*. American Media.

Utz, S., Wolfers, L. N., & Goritz, A. S. (2021). The effects of situational and individual factors on algorithm acceptance in COVID-19-related decision-making: A preregistered online experiment. *Human-Machine Communication*, *3*, 27–45. https://doi.org/10.30658/hmc.3.3

Walter, N., Cody, M. J., & Ball-Rokeach, S. J. (2018). The ebb and flow of communication research: Seven decades of publication trends and research priorities. *Journal of Communication*, *68*(2), 424–440. https://doi.org/10.1093/joc/jqx015

Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-it, I-thou, I-robot: The perceived humanness of AI in human-machine communication. *Communication Studies*, *71*(3), 393–408. https://doi.org/10.1080/10510974.2020.1749683

Woods, H. S. (2018). Asking more of Siri and Alexa: Feminine persona in service of surveillance capitalism. *Critical Studies in Media Communication*, *35*(4), 334–349. https://doi.org/10.1080/15295036.2018.1488082

Yan, H. Y., Yang, K. C., Menczer, F., & Shanahan, J. (2021). Asymmetrical perceptions of partisan political bots. *New Media & Society*, *23*(10), 3016–3037. https://doi.org/10.1177/1461444820942744

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

Zhu, Y., & Fu, K. W. (2019). The relationship between interdisciplinarity and journal impact factor in the field of communication during 1997–2016. *Journal of Communication*, *69*(3), 273–297. https://doi.org/10.1093/joc/jqz012

TABLE 1 Total Number of HMC Articles by Journal Ranked by Percentage of HMC Articles			
Journal	Amount of HMC Articles	Total Articles in Journal	Percentage of HMC Articles
Human-Machine Communication (HMC)	25	25	100%
Communication Studies (CS)	20	444	4.5%
Journal of Communication Pedagogy (JCP)	4	88	4.5%
Journal of Computer Mediated Communication (JCMC)	13	314	4.1%
Communication Research Reports (CRR)	13	447	2.9%
First Amendment Studies/Communication and Democracy (FAS)	2	117	1.7%
New Media & Society (NM&S)	21	1,320	1.6%
Communication Education (CE)	4	267	1.4%
Review of Communication (RC)	4	278	1.4%
Mobile Media & Communication (MM&C)	3	220	1.3%
Quarterly Journal of Speech (QJS)	3	222	1.3%
Critical Studies in Media Communication (CSMC)	4	342	1.1%
Communication Reports (CR)	1	132	0.75%
Communication, Culture and Critique (CCC)	3	431	0.69%
Communication Research (CRE)	3	490	0.61%
Communication and Critical/Cultural Studies (CCCS)	2	348	0.57%
Human Communication Research (HCR)	1	251	0.39%
Journal of Communication (JOC)	2	527	0.37%
Communication Quarterly (CQ)	1	348	0.28%
Asian Journal of Communication (AJC)	1	365	0.27%
Communication Theory (CT)	1	272	0.26%
Communication Teacher (CTE)	1	453	0.2%
Communication Monographs (CM)	0	268	0.0%
Journal of Applied Communication Research (JACR)	0	307	0.0%
Journal of International and Intercultural Communication (JIIC)	0	230	0.0%
Southern Communication Journal (SCJ)	0	291	0.0%
Text and Performance Quarterly (TPQ)*	0	326	0.0%
Western Journal of Communication (WJC)	0	374	0.0%
Total	132	9,497	1.39%

Note. For the year 2021 issues 3 and 4 of *Text and Performance Quarterly* were not completed at the time of analysis.

TABLE 2   HMC Articles by Year and Journal

	HMC	CS	CRR	SJC	CQ	CR	WJC	HCR	JCMC	CT	JOC	CE	JACR	CM	CRE	NM&S	AJC	MM&C	JCP	QJS	CCCS	CTE	CSMC	FAS	JIIC	TPQ	RC	CCC	Total
2011	Np	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	Np	0	0	1	0	0	0	0	0	0	3
2012	Np	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	Np	0	0	0	0	0	0	0	0	0	0
2013	Np	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	Np	0	0	0	0	0	0	0	0	0	1
2014	Np	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	Np	0	0	0	1	0	0	0	0	0	4
2015	Np	0	1	0	0	0	0	0	2	0	0	0	0	0	0	1	0	0	Np	0	0	0	0	0	0	0	0	0	4
2016	Np	2	1	0	0	0	0	0	0	0	0	0	0	0	2	1	0	0	Np	0	1	0	0	0	0	0	0	0	7
2017	Np	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	Np	1	0	0	0	0	0	0	1	0	4
2018	Np	0	5	0	0	0	0	0	0	0	1	1	0	0	0	3	0	0	0	0	0	0	2	1	0	0	1	0	14
2019	Np	1	1	0	0	0	0	0	1	0	0	0	0	0	0	3	0	2	0	1	0	0	1	0	0	0	0	2	12
2020	8	7	1	0	0	0	0	0	3	0	0	1	0	0	1	2	0	0	0	1	0	0	0	1	0	0	0	1	26
2021	17	10	3	0	1	1	0	0	4	1	0	2	0	0	0	10	1	0	4	0	1	0	0	0	0	0	2	0	57
Total	25	20	13	0	1	1	0	1	13	1	2	4	0	0	3	21	1	3	4	3	2	1	4	2	0	0*	4	3	132

Note. For the year 2021 issues 3 and 4 of *Text and Performance Quarterly* were not completed at the time of analysis. Np = Not published.

TABLE 3 Top 3 Most Cited Articles			
Author	**Year**	**Journal**	**Times Cited (Crossref)**
Ho, A., Hancock, J. T., & Miner, A. S.	2018	JOC	79
Guzman, A. L., & Lewis, S.C.	2020	NM&S	78
Spence, P. R., Westerman, D., Edwards, C., & Edwards, A.	2014	CRR	69

TABLE 4 Top 3 Articles in Altmetric Score			
Author	**Year**	**Journal**	**Altemtric Score**
Woods, H. S.	2018	CSMC	143
Guzman, A. L., & Lewis, S. C.	2020	NM&S	111
Yan, H. Y., Yang, K. C., Menczer, F., & Shanahan, J.	2021	NM&S	78

The Symptom of Ethics: Rethinking Ethics in the Face of the Machine

David J. Gunkel[1]

1 Department of Communication, Northern Illinois University, DeKalb, IL, USA

Abstract

This essay argues that it is the machine that constitutes the symptom of ethics—"symptom" understood as that excluded "part that has no part" in the system of moral consideration. Ethics, which has been historically organized around a human or at least biological subject, needs the machine to define the proper limits of the moral community even if it simultaneously excludes such mechanisms from any serious claim on moral consideration. The argument will proceed in five steps or movements. The first part will define and characterize "the symptom" as it has been operationalized in the work of Slovenian philosopher Slavoj Žižek. Although Žižek appropriates this term from Jacques Lacan, he develops the concept in a unique way that exceeds Lacan's initial psychoanalytic formulations. The second and third parts will demonstrate how the machine constitutes the symptom of moral philosophy, showing how and why it comprises the always already excluded element necessary to define the proper limits of moral subjectivity. The fourth part will then consider two alternatives that promise, but ultimately fail, to accommodate this symptom. And the final section will draw out the consequences of this analysis for ethics and its excluded others.

Keywords: artificial intelligence, ethics, machine, psychoanalysis, robot

One of the persistent challenges of moral decision-making is determining exactly who or what deserves ethical consideration. Although initially limited to "other men," the practice of ethics has evolved in such a way that it continually challenges its own exclusions and

comes to encompass previously marginalized individuals and groups—foreigners, women, animals, and even the environment. Despite these progressive efforts at inclusion or what animal rights philosopher Peter Singer (1973) has called "liberation movements," one thing remains outside the community of legitimate social subjects—the machine. Traditionally characterized as a mere instrument or "means to an end," these technological artifacts have been and remain the excluded other. As J. Storrs Hall (2001) explains, "we have never considered ourselves to have 'moral' duties to our machines, or them to us" (unpaginated). And yet, despite this almost absolute categorical exclusion, ethics seems to need and to be unable to do without these mechanisms.

This essay argues that it is the machine that constitutes the symptom of ethics—"symptom" understood as that excluded "part that has no part." Ethics, which has been historically organized around a human or at least biological subject, needs the machine to define the proper limits of the moral community even if it simultaneously excludes such mechanisms from any serious claim on ethics. Consequently, the machine in general and the automaton, or self-moving machine in particular, constitutes the other of the Other that must be excluded in order to define the legitimate boundaries of the moral community (e.g., who is and what is not considered to be a subject of rights and obligations).

The argument will proceed in several steps or movements. The first will define and characterize "the symptom" as it has been operationalized in the work of Slavoj Žižek. Although Žižek appropriates this term from Jacques Lacan, he develops the concept in a unique way that exceeds Lacan's initial psychoanalytic formulations. The second and third parts will demonstrate how the machine constitutes the symptom of ethics, showing how and why it comprises the always already excluded element necessary to define the proper limits of moral subjectivity. The fourth part will then consider two alternatives that promise, but ultimately fail, to accommodate this symptom. And the final section will draw out the consequences of this analysis for communication ethics and the excluded other.

A Symptom of the Symptom

Slavoj Žižek has been concerned with the concept of the symptom from the very beginning. In fact, "The Symptom" comprises the subject matter of the first part of his first book published in English, *The Sublime Object of Ideology* (first published in 1989, second edition issued in 2008). This text begins in a way that is rather characteristic of all Žižek's writings—with a statement from Jacques Lacan that seems, at first, to be counterintuitive: "According to Lacan, it was none other than Karl Marx who invented the notion of symptom" (Žižek, 1989/2008, p. 3). And the two chapters that follow this statement, "How Marx Invented the Symptom" and "From Symptom to Sinthome," are designed to explicate and develop this insight. In doing so, Žižek provides a characterization of the symptom that can itself be considered symptomatic of Western metaphysics.

The word "symptom" is typically used to indicate a mode of indication. As Todd McGowan (2014) explains for the entry "symptom" in the *Žižek Dictionary*, "the usual idea of the symptom in both psychoanalysis and traditional medicine sees it as an indication of an underlying disorder that some form of therapy (either analytic or medicinal) will attempt to cure and thereby eliminate" (p. 242). Formulated in this way, "symptom" is understood as

a sign that refers to and is the external manifestation or indication of something that is more fundamental and often hidden from direct perception. We therefore commonly distinguish the symptom from its underlying cause, and this common understanding not only adheres to the standard formula of semiotics but is informed by an original ontological decision that differentiates between mere external appearances and the more profound substance that is its ultimate cause and referent. Characterized in this fashion, the usual understanding of "symptom" can be accommodated to and explained by a metaphysical arrangement that is at least as old as Plato. According to the standard account of Platonism, or what is often called (not without controversy) "Plato's theory of the forms," appearances are nothing less than symptoms of more substantial transcendental ideas, and the task of thinking is to learn to penetrate or see beyond these mere external apparitions and gain access to the true form and original cause.

For Žižek, however, the symptom is formulated in a way that is entirely otherwise. "Symptom" is not the sign of some hidden kernel of truth that is more substantial or profound; it is the necessarily excluded other of the system that makes the system possible in the first place. As Žižek (1989/2008) explains,

> the 'symptom' is, strictly speaking, a particular element which subverts its own universal foundation, a species subverting its own genus . . . a point of breakdown heterogeneous to a given ideological field and at the same time necessary for that field to achieve its closure, its accomplished form. (p. 16)

For this reason, "symptom" possesses a "radical ontological status" (Žižek, 1989/2008, p. 81); it is the constitutive part of a system that is necessarily excluded from the system as such. Or as Žižek (1999) explains by way of a passage he appropriates from Jacques Rancière, it is "the part of no part" (p. 188). The symptom, therefore, comprises the constitutive exception that "threatens the functioning of the system, even though it is the necessary product of this same system" (McGowan, 2014, p. 242). It is a kind of unacknowledged (and always already unacknowledgeable) excremental reminder that is necessary for something to produce itself and function as the system that it is.

This alternative conceptualization of the symptom—a conceptualization that is simultaneously dependent on Lacan's work and beyond the circuit of its determinations—turns out to be symptomatic of Platonism. And the fact that Žižek himself never actually acknowledges this as such is just one more symptom of the symptom. It is, in fact, only by excluding this particular concept of the symptom (the very idea of the "constitutive exception"), that Platonism can become what it is. One might recall that Socrates, as was described in Plato's *Apology*, explains and tries to defend his own efforts as a response to the Oracle at Delphi (Plato, 1982, p. 21a). The Delphic temple, as is reported in *Protagoras* (Plato, 1977, p. 324b), famously had two laconic statements inscribed above its gate: "Know thyself" and "Nothing in Excess." Although the latter has typically been interpreted as a call to moderation in all things, it can also be read as the trace of the symptom, inscribed at the very gateway to knowledge. In this way, the statement "Nothing in excess" may be interpreted as an exclusive operation, indicating that whatever might come to exceed the grasp of self-knowing is to remain unacknowledged, unknowable, or nothing. Consequently, every attempt to

demarcate the proper boundaries of a system and cordon off its internal workings from what it is not, always and without exception, produces an exceptional externality that it must deny—an absolutely other that is cast off, externalized, and remains nothing.

The Symptom of Ethics

If Žižek is correct, then the system of ethics must have its symptom, or better, its *sinthome*. For according to Žižek, "each subject or system has one fundamental symptom that contains the key to its structure and holds this structure together. This fundamental symptom is what Lacan calls the *sinthome*" (McGowan, 2014, pp. 242–243). And the *sinthome*—or the "constitutive exception," if one prefers to avoid psychoanalytic "jargon"—of the system of ethics will have been the machine, and it "will have been" because the symptom arrives not from the depths of the past but from the future (but more on this later, that is, in the future.).

It was, we might say parroting Žižek, none other than René Descartes who invented the symptom of modern science. This is because Descartes, in the course of developing his particular brand of philosophical self-reflection, cordoned off the human from its others, specifically the animal and the machine. In fact, Descartes executes this exclusion by associating the animal with the machine, introducing an influential concept—the doctrine of the *bête-machine* or *animal-machine*. "Perhaps the most notorious of the dualistic thinkers," Akira Mizuta Lippit (2000) writes,

> Descartes has come to stand for the insistent segregation of the human and animal worlds in philosophy. Likening animals to automata, Descartes argues in the 1637 *Discourse on the Method* that not only 'do the beasts have less reason than men, but they have no reason at all.' (p. 33)

For Descartes, the human being was considered the sole creature capable of rational thought—the one entity able to say and be certain in its saying, *cogito ergo sum*. Following from this, he had concluded that other nonhuman animals not only lacked reason but were nothing less than mindless automata that, like clockwork mechanisms, simply followed pre-programmed instructions.

Conceptualized in this fashion, the animal and machine were effectively indistinguishable and ontologically the same. "If any such machine," Descartes (1988) wrote, "had the organs and outward shape of a monkey or of some other animal that lacks reason, we should have no means of knowing that they did not possess entirely the same nature as these animals" (p. 44). Beginning with Descartes, then, the animal and machine share a common form of alterity that situates them as completely different from and distinctly other than human. [1] Despite pursuing a method of doubt that, as Jacques Derrida (2008) describes it, reaches "a level of hyperbole," Descartes "never doubted that the animal was only a machine" (p. 75). The *bête-machine*, we can say following Žižek, is the symptom of Cartesian thought. It is the necessarily excluded other of human rationality, and it is only through this mechanism of exclusion and its exclusivity that the system of human rationality (and rationality as specifically defined by an act of self-knowing) emerges and functions.

Following Descartes's exclusive ontological decision, animals have not traditionally been considered a legitimate subject of moral consideration. Determined to be mere mechanisms, they were nothing more than instruments to be used more or less effectively by human beings who were typically the only subjects who mattered. When Immanuel Kant (1985), for instance, defined morality as involving the rational determination of the will, the animal, which does not by definition possess reason, is immediately and categorically excluded. [2] The practical employment of reason does not concern the animal and, when Kant does make mention of animality [*Tierheit*], he only does so in order to use it as a foil by which to define the limits of humanity proper. The same exclusions have been instituted and supported by subsequent works. According to Tom Regan, this decision also affects and is apparent in the seminal work of analytic ethics.

> It was in 1903 when analytic philosophy's patron saint, George Edward Moore, published his classic, *Principia Ethica*. You can read every word in it. You can read between every line of it. Look where you will, you will not find the slightest hint of attention to 'the animal question.' Natural and nonnatural properties, yes. Definitions and analyses, yes. The open-question argument and the method of isolation, yes. But so much as a word about non-human animals? No. Serious moral philosophy, of the analytic variety, back then did not traffic with such ideas. (Regan, 1999, p. xii)

Consequently, ethics, like any closed system, must have and cannot do without its symptoms—the constitutive exception that although excluded from participation, necessarily make it what it is. This exclusivity is fundamental, structural, and systemic. It is not accidental, contingent, or prejudicial in the usual sense of those words. "When it comes to moral considerability," Thomas Birch (1993) explains, "there are, and ought to be, insiders and outsiders, citizens and non-citizens . . . 'members of the club' of *consideranda* versus the rest" (p. 315). And because this exclusivity is systemic, little or nothing changes as moral theory and practice has developed and matured over time. Even when membership in the exclusive club of moral subjects has, slowly and not without considerable resistance and struggle, been extended to previously excluded others, there have remained other, apparently more fundamental and necessary exclusions. Or to put it another way, every new seemingly progressive inclusion has been made at the expense of others, who are necessarily excluded in the process.

Take, for example, innovations in animal rights. This rather recent development in moral thinking, not only challenged the anthropocentric tradition in ethics but redefined the club of *consideranda* by taking a distinctly animo-centric approach, where the qualifying criteria for inclusion in the community of moral subjects was not determined by some list of indeterminate humanlike capabilities—consciousness, rationality, free will, and so on—but sentience and the capacity to suffer. Although Jeremy Bentham is often credited with introducing the innovation, the movement does not coalesce until the later part of the 20th century. Tom Regan (1999) identifies the crucial turning point in a single work: "In 1971, three Oxford philosophers—Roslind and Stanley Godlovitch, and John Harris—published *Animals, Men and Morals*. The volume marked the first time philosophers had collaborated to craft a book that dealt with the moral status of nonhuman animals" (p. xi).

According to Regan, this particular publication is not only credited with introducing what is now called the "animal question," but launched an entire subdiscipline of ethics where the animal (or at least some order of animal) is considered to be a legitimate subject of moral inquiry.

What is remarkable about this development is that at a time when this other form of nonhuman otherness is increasingly recognized as a legitimate moral subject, its other, the machine, remains conspicuously absent and marginalized. Despite all the ink that has been spilled on the animal question, little or nothing has been written about the machine. One could, in fact, redeploy Regan's critique of G. E. Moore's *Principia Ethica* (2005) and apply it, with a high degree of accuracy, to any work purporting to address the animal question: "You can read every word in it. You can read between every line of it. Look where you will, you will not find the slightest hint of attention to 'the machine question.'" Even though the fate of the machine, from Descartes forward, was intimately coupled with that of the animal, only one of the pair has qualified for any level of ethical consideration. [3] Technology, as Jean-François Lyotard (1993, p. 44) reminds us, is only a matter of efficiency. Technical devices do not participate in the big questions of metaphysics, aesthetics, or ethics. They are nothing more than contrivances or extensions of human agency, used more or less responsibly by human agents with the outcome affecting other human individuals. Consequently, machines like computers, robots, and other kinds of mechanisms do not, at least for the majority philosophical opinion, have an appropriate place within ethics. Although other kinds of previously excluded others have been slowly and not without struggle been granted membership in the community of moral subjects—women, people of color, some animals, and even the environment—the machine remains on the periphery. No matter how automatic, interactive, or intelligent it might appear to be, the machine remains the excluded other of the other—the constitutive exception of moral philosophy's increasing attention to previously excluded others.

Enjoy Your Symptom

Despite its exclusion from consideration, the machine has been and remains fundamental to the system of ethics. In other words, moral theory and practice, although excluding the machine from explicit consideration needs and depends on it for its own systematicity. All too often, however, one misses this fact, because of the way we have (mis)understood and restricted the definition of the machine. Typically, the word "machine" is understood and characterized as a physical mechanism. "We have a naïve notion of a machine as a box with motors, gears, and whatnot in it" (Hall, 2001, unpaginated). It is, for example, the spring-driven mechanical clock, introduced in Europe around the middle of the 16th century, which had comprised the principal machinic prototype for much of the modern period. For Descartes, the mechanical clock, with its intricate gears, was a model of the mindless animal body, which moves itself and responds to stimulus like a well-fashioned mechanism. In Sir Isaac Newton's *Philosophiae Naturalis Principia Mathematica*, this image was extended to cover the entirety of physical reality, introducing a concept that has come to be called the "clockwork universe" (Newton, 1972). Even after technology had advanced well beyond the gears, springs, and levers of the clock, philosophers continued to fixate on mechanics. For Martin Heidegger (1977), for example, technology was restricted to mechanical

apparatuses: sawmills, hydroelectric power plants, high-frequency radar stations, and jet aircraft (p. 5).

This particular definition of the machine is not only restrictive but, as Hall (2001) argues, "incapable of dealing with the machines of the future" (unpaginated). According to Hall, a machine is not simply a combination of gears and motors. It is a set of rules, instructions, or messages.

> The most important machine of the twentieth century wasn't a physical thing at all. It was the Turing Machine, and it was a mathematical idea. It provided the theoretical basis for computers . . . This theoretical concept of a machine as a pattern of operations which could be implemented in a number of ways is called a virtual machine. (Hall, 2001, unpaginated)

Understood in this fashion, "machine" is not merely a collection of physical springs and gears but a system of encoded instructions, an algorithm, which may be implemented and embodied in any number of ways. This general definition of the machine covers mechanical systems, like clocks that implement rules of synchronization in the form of physical space marked out on rotating gears; biological systems, like animals and plants that are composed of and operate by following instructions embedded in their genetic code; and information processing devices, like the computer, which performs different operations based on various program instructions stipulated by software. As Donna Haraway (1991) has argued, following the innovations introduced by Norbert Wiener, we are all understood and constituted as mechanisms of communication.

If the machine, according to this general definition, is a pattern of operations or a set of pre-defined instructions, then ethics has been and continues to be mechanistic. According to the English moralist Henry Sidgwick (1981), "the aim of Ethics is *to systematize* and free from error the apparent cognitions that most men have of the rightness or reasonableness of conduct" (p. 77). Western conceptions of morality customarily consist in systematic rules of behavior that can be encoded, like an algorithm, and implemented by different moral agents in a number of circumstances and situations. They are, in short, an instruction set that is designed to direct behavior and govern conduct. Take for example, the Ten Commandments, the cornerstone of Judeo-Christian ethics. These ten rules constitute something of a moral subroutine that not only prescribe correct operations for human beings but do so in a way that is abstracted from the particulars of circumstance, personality, and other empirical accidents. "Thou shall not kill" is a general prohibition against murder that applies in any number of situations where one human being confronts another. Like an algorithm, the statements contained within the Ten Commandments are general operations that can be applied to any particular set of data.

Similarly, Immanuel Kant's moral philosophy is founded on and structured by fundamental rules or what he calls, in a comparison to natural science, "practical laws." These practical laws are "categorical imperatives." That is, they are not merely subjective maxims that apply to a particular person's will under a specific set of circumstances. Instead, they must be universally and objectively valid for the will of every rational being in every possible circumstance.

> Laws must completely determine the will as will, even before I ask whether I am capable of achieving a desired effect or what should be done to realize it. They must thus be categorical; otherwise they would not be laws, for they would lack the necessity which, in order to be practical, must be completely independent of pathological conditions, i.e., conditions only contingently related to the will. (Kant, 1985, p. 18)

For Kant, moral action is programmed by principles of pure practical reason—universal laws that are not only abstracted from every empirical condition but applicable to any and all rational agents. It may be said, therefore, that Kant, who took physics and mathematics as the model for a wholesale transformation of the procedures of philosophy, mechanized ethics in a way that was similar to Newton's mechanization of physical science.

Finally, even the pragmatic alternative to deontological ethics, utilitarianism, operates by a kind of systemic moral computation or what Jeremy Bentham (1988) called "moral arithmetic." The core utilitarian principle, "seek to act in such a way as to promote the greatest quantity and quality of happiness for the greatest number," is a general formula that subsequently requires considerable processing to crunch the numbers and decide the best possible outcome. In fact, Michael Anderson, Susan Leigh Anderson, and Chris Armen (2004) have not only constructed computer-based "ethical advisors" but argue that such machines might have an advantage over a human being in following utilitarian theory, because of the sheer number of variables that usually need to be taken into account and calculated accurately (p. 2). For this reason, Joseph Nadeau (2006) has even suggested that machines might be more moral, that is, less biased and more reasonable in executing moral decision-making. In all these cases, ethics—which, according to Sidgwick (1981, p. 77), aims to systematize human cognition and conduct—conforms to the characterization of what is called "the virtual machine." Commandments, moral imperatives, ethical principles, codes of conduct, practical laws . . . these all endeavor to provide a clear set of instructions or patterns of operation that are designed to program and direct human social behavior and interaction.

How to Survive the Robot Apocalypse

The system of ethics is both dependent upon and exclusive of the machine. It comprises that necessary part of moral thinking that morality wants no part of. And it is for this reason that it initially appears to us as a kind of external or alien threat. In fact, in the popular imagination, the machine—a generic name that is, following Derrida (2008), admittedly as problematic as the term "animal" (p. 23)—is often imagined as coming at us from another time and place. The "machinic other" is typically portrayed in the form of an invading army of robots descending on us from the outer reaches of space sometime in the not-too-distant future. This particular representation is not accidental or a mere artifact of contemporary popular culture. There is a good reason for it. As Žižek (1989/2008) explains "symptoms are meaningless traces, their meaning is not discovered, excavated from the hidden depths of the past, but constructed retroactively" from the future (p. 58). If the moral challenge of the machine is typically imagined in terms of this popular sci-fi formula, it is because it constitutes the constitutive exception of moral philosophy and can appear to us—assuming that

it does appear to us—only as a kind of external threat proceeding from the future. In this, Deleuze (1994) was undeniably correct—philosophy is a kind of science fiction (p. xx). The question then, is how do or should we respond in the face of this apparent robot invasion? There are at least two alternatives situated on opposite ends of the philosophical spectrum, neither of which is entirely adequate.

Moral Totalitarianism

One method of response, what we might call "moral totalitarianism," is to develop a more inclusive and complete moral system that can incorporate not just the machine but any and all symptomatic others. Take, for example, Luciano Floridi's (2013) proposal for a new moral theory he calls "Information Ethics," abbreviated IE. According to Floridi, efforts to evolve moral consideration have been woefully inadequate insofar as each new innovation cannot, it seems, succeed without making additional exclusions. Although he does not say it in this exact way, what concerns Floridi is the symptomatic remainder of ethics. Animal rights philosophy, he points out, correctly challenged the anthropocentric tradition. But it succeeded only by further excluding other living organisms, like plants and the environment. As a result, the innovations of bioethics and environmental ethics sought to repair this exclusivity by developing a moral theory that was more inclusive of these previously excluded others. But, Floridi (2013) continues, even the innovations of

> bioethics and environmental ethics fail to achieve a level of complete universality and impartiality, because they are still biased against what is inanimate, lifeless, intangible, abstract, engineered, artificial, synthetic, hybrid, or merely possible. Even land ethics is biased against technology and artefacts, for example. From their perspective, only what is intuitively alive deserves to be considered as a proper centre of moral claims, no matter how minimal, so a whole universe escapes their attention. (p. 64)

IE is designed to respond to this exclusivity by developing,

> an ecological ethics that replaces biocentrism with ontocentrism. IE suggests that there is something even more elemental than life, namely being—that is, the existence and flourishing of all entities and their global environment—and something more fundamental than suffering, namely entropy, [which] here refers to any kind of destruction or corruption of informational objects, that is, any form of impoverishment of being including nothingness, to phrase it more metaphysically. (Floridi, 2008, p. 47)

Following the moral innovations of bio- and environmental ethics, Floridi advocates expanding the scope of ethics by altering its focus and lowering the threshold for inclusion, or, to use Floridi's terminology, the "level of abstraction" (LoA). What makes someone or something a moral subject, deserving of some level of ethical consideration, is that it exists as a coherent body of information. "Unlike other non-standard ethics," Floridi (2013) argues that,

> IE is more impartial and universal—or one may say less ethically biased—
> because it brings to ultimate completion the process of enlarging the concept of
> what may count as a centre of moral claims, which now includes every instance
> of information, no matter whether physically implemented or not. (p. 65)

The proposal certainly sounds promising. Like previous innovations and "liberation movements" (Singer, 1973) in ethics, IE is interested in expanding membership in the moral community so as to incorporate previously excluded others and eliminate the symptom of ethics. But, unlike these previous efforts, it is arguably more inclusive, incorporating other forms of otherness, like technologies, artifacts, and machines. But IE, for all its promise to provide what one might call "a moral theory of everything," still makes exclusive decisions. As a system of ethics it too must have and cannot proceed without its symptom. And what is excluded from this seemingly complete and all-encompassing moral theory is "nothing." [4] As if following the two laconic imperatives inscribed above the gate at Delphi, IE's total-izing comprehension leaves only "nothing in excess." But this "nothing" is not no-thing; it is IE's particular symptom. Consequently, the problem with efforts at fabricating increasing greater levels of moral inclusivity, like that proposed by IE, is that they remain symptomatic.

Thinking Otherwise

The alternative to this moral totalitarianism proceeds and operates otherwise. And when it comes to thinking otherwise, especially as it relates to questions regarding ethics, there is perhaps no thinker better suited to the task than Emmanuel Levinas. Unlike a lot of what goes by the name of "moral philosophy," Levinasian thought does not rely on metaphysical generalizations, abstract formulas, or simple pieties. Levinas (1969) is not only critical of the traditional tropes and traps of Western ontology but proposes an ethics of radical other-ness that deliberately resists and interrupts the metaphysical gesture par excellence, that is, the reduction of difference to the same. This radically different approach to thinking differ-ence differently is not a gimmick. It constitutes a fundamental reorientation that effectively changes the rules of the game and the standard operating presumptions.

Levinas, therefore, deliberately turns things around by reconfiguring the assumed order of precedence in situations regarding moral decision-making. For him, the ethical relation-ship, the moral obligation that I have to the Other, precedes and determines who or what comes, after the fact, to be considered a moral subject or "person." [5] Ethics, therefore, is not predicated on an *a priori* determination of who or what is a legitimate moral subject. Instead, moral standing is something that is first decided on the basis of and as a product of a social encounter. According to Levinas, therefore, the Other always and already obligates me in advance of the customary decisions and debates concerning who or what is and is not considered a moral subject. This apparent inability or indecision is not necessarily a prob-lem. In fact, it is a considerable advantage insofar as it opens ethics not only to the Other but to other forms of otherness. "If this is indeed the case," as Matthew Calarco (2008) concludes,

> that is, if it is the case that we do not know where the face begins and ends, where
> moral considerability begins and ends, then we are obligated to proceed from

the possibility that anything might take on a face. And we are further obligated to hold this possibility permanently open. (p. 71)

Despite the promise this innovation has for arranging an ethics that is oriented otherwise, Levinas's work is not, it seems, able to escape from the traditional anthropocentric privilege. Whatever the import of his unique contribution, Other in Levinas is still (for better or worse) unapologetically human. If, according to Levinas, previous forms of moral theorizing can be criticized for putting ontology before ethics, then Levinasian thought can be cited for its unquestioned philosophical anthropology. If Levinasian thought is to provide a way of thinking that is able to respond to and to take responsibility for these other forms of otherness, or to consider and respond to, as John Sallis (2010) describes it, "the question of another alterity" (p. 88), we will need to use and interpret Levinas's own philosophical innovations in excess of and in opposition to him. We will need, as Derrida (1978) once wrote of Georges Bataille's exceedingly careful engagement with the thought of Hegel, to follow Levinas to the end, "to the point of agreeing with him against himself" (p. 260) and of wresting his discoveries from the limited interpretations that he provided. Such efforts at "radicalizing Levinas" (Atterton & Calarco, 2010) will take up and pursue Levinas's moral innovations in excess of the rather restricted formulations that he and his advocates and critics have typically provided. "Although Levinas himself," Calarco (2008) writes,

> is for the most part unabashedly and dogmatically anthropocentric, the underlying logic of his thought permits no such anthropocentrism. . . . In fact, as I shall argue, Levinas's ethical philosophy is, or at least should be, committed to a notion of universal ethical consideration, that is, an agnostic form of ethical consideration that has no a priori constraints or boundaries. (p. 55)

This radical reorientation—or other version of Levinas's ethics of otherness—obviously opens the door to what some might consider absurd possibilities:

> At this point, most reasonable readers will likely see the argument I have been making as having absurd consequences. While it might not be unreasonable to consider the possibility that 'higher' animals who are 'like' us, animals who have sophisticated cognitive and emotive functions, could have a moral claim on us, are we also to believe that 'lower' animals, insects, dirt, hair, fingernails, ecosystems and so on could have a claim on us? (Calarco, 2008, p. 71)

In responding to this charge, Calarco deploys that distinctly Žižekian strategy of "fully endorsing what one is accused of" (Žižek, 2000, p. 2). "I would suggest," Calarco (2008) argues,

> affirming and embracing what the critic sees as an absurdity. All attempts to shift or enlarge the scope of moral consideration are initially met with the same reactionary rejoinder of absurdity from those who uphold common sense. But any thought worthy of the name, especially any thought of ethics, takes its point of departure in setting up a critical relation to common sense and the established doxa and, as such, demands that we ponder absurd, unheard-of thoughts. (p. 72)

Calarco's reworking of Levinasian philosophy seems to provide for a more inclusive ethics that is able to take other forms of otherness into account. And it is, no doubt, a compelling proposal. What is interesting about his argument, however, is not the other forms of otherness that come to be incorporated through his innovative reworking of Levinas, but what gets left out in the process. For all its promise to think ethics otherwise, Calarco's radicalization of Levinas still has its symptoms. According to the letter of Calarco's text, the following entities should also be included as potentially significant: "'lower' animals, insects, dirt, hair, fingernails, and ecosystems." What is obviously missing from this list is anything that is not "natural," that is, any form of artifact. Consequently, what gets left out by or excluded from Calarco's "universal consideration"—a mode of ethical concern that does not shrink from potential absurdities and the unthinkable—are tools, technologies, and machines. Although Calarco (2008) is clearly prepared, in the name of the other and other kinds of otherness, "to ponder absurd, unheard-of thoughts" (p. 72) the machine remains excluded and in excess of this effort, comprising a kind of absurdity beyond absurdity, the unthinkable of the unthought, or the other of all who are considered Other. The alterity of all kinds of other nonhuman things does, in fact, and counter to Levinas's own interpretation of things, make an ethical impact. But this does not apply, it seems, to machines, which remain, for Levinas, Calarco, and others, the excluded other of moral philosophy's own interest in otherness.

Summary and Conclusions

"Every philosophy," Silva Benso (2000) writes in a comprehensive gesture that performs precisely what it seeks to address, "is a quest for wholeness." This objective, she argues, has been typically targeted in one of two ways. "Traditional Western thought has pursued wholeness by means of reduction, integration, systematization of all its parts. Totality has replaced wholeness, and the result is totalitarianism from which what is truly other escapes, revealing the deficiencies and fallacies of the attempted system" (p. 136). This is precisely the problem with totalizing systems of moral inclusion, like Floridi's Information Ethics. For all its efforts at achieving a more inclusive form of inclusion, IE still makes a distinction between inside and outside—between who matters and what does not—even if the symptom of IE is nothing. [6]

The competing alternative to this totalitarian approach is a philosophy of difference that is oriented otherwise, like that proposed and developed by Levinas, Calarco, and others. This other approach endeavors to achieve moral completion,

> by moving not from the same, but from the other, and not only the Other, but also the other of the Other, and, if that is the case, the other of the other of the Other. In this must, it must also be aware of the inescapable injustice embedded in any formulation of the other. (Benso, 2000, p. 136)

For Levinas and those others who endeavor to develop this particular brand of thinking otherwise, every other has its other such that the process of responding to previously excluded others is never fully complete. What is interesting about these two strategies is not what makes them different from one another or how they articulate approaches that

proceed from what appears to be opposite ends of the spectrum. What is interesting is what they agree on and hold in common in order to be situated as different from and in opposition to each other in the first place.

Whether taking the form of a totalizing autology or an alternative kind of heterology, both approaches "share the same claim to inclusiveness" (Benso, 2000, p. 136), and that is the problem. We therefore appear to be caught between a proverbial rock and a hard place. On the one hand, inclusion has never been inclusive enough. The machine in particular is from the very beginning situated outside ethics. It is, irrespective of the different philosophical perspectives that come to be mobilized, not a legitimate moral subject. And even when, at the apparent apex of moral inclusivity with the innovative efforts of IE, the machine can be accommodated, this inclusion cannot succeed apart from instituting additional exclusions and marginalizations. On the other hand, alternatives to this tradition have never quite been different enough. Although a concern with and for others promises to transform the status quo in ethics, "thinking otherwise" has never been entirely adequate or suitably different. Many of the so-called alternatives, those efforts that purport to be interested in and oriented otherwise, have typically excluded the machine from the space of difference, from the difference of difference, or from the otherness of the Other. Technological devices certainly have an interface, but they do not it seems, possess a face or confront the human user in a face-to-face encounter that would call for and would be called ethics.

The problem with both approaches is that they seek a utopian outcome. As Žižek (1989/2008) explains, "'utopian' conveys the belief in the possibility of a universality without its symptom, without the point of exception functioning as its internal negation" (p. 13). This utopianism, however, never succeeds. Each innovative effort at moral inclusion produces a remainder. Each new system of ethics cannot help but generate its symptom. Or as McGowan (2014) explains in a more politically situated context, "one cannot simply expand representation to include them because some new excluded group will always come to occupy this position" (p. 243). This is because the mechanism of exclusion is systemic and has little or nothing to do with the actual "things" that are subjected to marginalization. The exclusivity of the machine, therefore, is not simply "the last socially accepted prejudice" or what Singer (1989) calls "the last remaining form of discrimination" (p. 148), which may be identified as such only from a perspective that is already open to the possibility of some future inclusion and accommodation. It is systemic and comprises the symptom of ethics.

Although Žižek does not necessarily provide a definitive solution to this impasse, he does indicate what would be necessary for an alternative kind of *eccentric* moral theory, called this because it would be an ethics without a clearly defined "center" as has been the case for other moral theories like anthropocentrism, animocentrism, biocentrism, and ontocentrism. Unlike Floridi, Levinas, and others, Žižek does not play the game of trying to remediate the symptom of ethics by designing systems for greater inclusivity. Instead, he proposes *an ethics of the symptom*, which would be not an(other) exclusive moral theory but a moral philosophy of the excluded. He therefore proposes a community of moral subjects consisting of nothing but a loose amalgam of excluded misfits, or what Alphonso Lingis (1994) calls "the community of those who have nothing in common." Though this proposal can also be called "utopian," it is a significantly different and somewhat distorted form of utopia. Žižek, as McGowan (2014) explains, does not

dismiss out of hand all utopian thinking. In fact, he constructs a utopianism based on the symptom, a utopianism in which a community forms from the exclude rather than through a universal inclusion. All those who exist outside the system as its symptoms can come together in a universal solidarity. This solidarity would not involve any sense of belonging because what the subjects have in common is only their exclusions or symptomatic status. (pp. 243–244)

What Žižek proposes, therefore, would be an eccentric community of eclectic elements that does not simply oppose one form or method of inclusivity with another, seemingly more inclusive form—the way that, for example, the ontocentrism of IE challenges the exclusions of biocentrism or Calarco's "radicalizing Levinas" questions the exclusivity of Levinas's philosophical anthropology. This eccentric form of moral thinking recognizes that the real challenge for ethics is not figuring out a way to include others, but to identify and confront the systemic exclusions of any and all efforts at inclusion as a significant and fundamental aspect of moral thinking itself. What we need to do in the face of the machine, therefore, is not to try to formulate more inclusive forms of moral theory that can account for and incorporate these others, but to recognize the symptom as such and allow it to question the entire history of ethics and its necessary and unavoidable exclusions. This is precisely that kind of thinking that Friedrich Nietzsche (1966) had called "the philosophy of the future," not only because the symptom of ethics, like the machine, appears to threaten us from the future but because it points in the direction of a kind of thinking that is situated beyond (the very system of) good and evil. This means that the challenge presented to us by the machine is not just a matter of applied ethics; it invites and entrains us to rethink the entire *modus operandi* of moral philosophy all the way down. This is the task for thinking that is seen in the face or the faceplate of the machine [7].

Notes

[1] This modern innovation is also significant because it marked an important departure from medieval practices whereby animals were thought to be capable of committing crimes against human beings and put on trial for their transgressions. For more on this subject, see Beirnes (1994), Chesterman (2021), Evans (1906), and Kadri (2007).

[2] The decision to focus on *reason* as a qualifying criterion for inclusion in the community of moral subjects already hints at a potential problem and prejudicial exclusion. It is human beings—those entities who have defined themselves as *animal rationale*—who have decided that rationality (e.g., their own defining feature) is the exclusive qualifying characteristic. Human beings, therefore, grant to themselves the power and the privilege to be both the measure and measurer in matters concerning moral status. For more on this problem and its consequences, see Gunkel (2012).

[3] For a more detailed investigation of points of intersection and divergence between the "animal question" and the "machine question," see Gellers (2021), Gunkel (2012), and Hogan (2017).

[4] One might ask (as was the case with one of the reviewers of this essay): Why is the exclusion of nothing ethically problematic? It is a very reasonable question and one that appears to exonerate IE insofar as it could be said that this moral system is so complete in its efforts at inclusivity that it excludes nothing. Although an accurate statement, this is not what is of principal importance. What is important is that IE, like all other moral philosophies, still cannot do without or escape from its symptom (e.g., the necessary and systemic exclusions that are its condition of possibility). Because IE aims to be absolutely totalizing and inclusive of all being what remains excluded can only be nothing. This "nothing," however, is not no-thing. It is the symptom of IE. For a more complete analysis, see Gunkel (2012).

[5] This shift in focus has instituted what Coeckelbergh (2012) has called a relational turn in ethics. Other formulations of a relational approach to moral status ascription can be found in Abate (2019), Fox (1995), and Philippopoulos-Mihalopoulos (2017).

[6] Following this, we can say that the goal of justice is not and cannot be organized around efforts to develop a more inclusive (or totalizing) ethics by eliminating (or pretending to have eliminated) all that would have been excluded. Ethics is and cannot do without its symptom. The question is not whether there is an exclusive remainder or not. Instead, what matters is how a particular formulation of ethics responds to and takes responsibility for its own necessary and unavoidable systemic exclusions. This is the task not of ethics per se but of what Derrida (1978, p. 111) called "the ethics of ethics."

[7] My own vision of ethics following from this line of reasoning has been developed and presented in the books *The Machine Question* (Gunkel, 2012), *Robot Rights* (Gunkel, 2018), and *How to Survive a Robot Invasion* (Gunkel, 2020).

Author Biography

David J. Gunkel is an award-winning educator and scholar, specializing in the philosophy of technology with a focus on the ethics of emerging technology. He is the author of over 90 scholarly articles and has published thirteen books, including *Thinking Otherwise: Philosophy, Communication, Technology* (Purdue University Press 2007), *The Machine Question: Critical Perspectives on AI, Robots, and Ethics* (MIT Press 2012), *Robot Rights* (MIT Press 2018), and *An Introduction to Communication and Artificial Intelligence* (Polity 2020). He currently holds the position of Presidential Research, Scholarship and Artistry Professor in the Department of Communication at Northern Illinois University (USA). http://gunkelweb.com

https://orcid.org/0000-0002-9385-4536

References

Abate, R. S. (2019). *Climate change and the voiceless: Protecting future generations, wildlife and natural resources*. Cambridge University Press.

Anderson, M., Anderson, S. L., & Armen, C. (2004). Toward machine ethics. Paper presented at American Association for Artificial Intelligence—The Nineteenth National Conference on Artificial Intelligence, 25–29 July, San Jose, California. https://www.aaai.org/Papers/Workshops/2004/WS-04-02/WS04-02-008.pdf

Atterton, P., & Calarco, M. (2010) *Radicalizing Levinas*. SUNY Press.

Beirnes, P. (1994). The law is an ass: Reading E.P. Evans' *The Medieval Prosecution and Capital Punishment of Animals*. *Society and Animals, 2*(1), 27–46. https://doi.org/10.1163/156853094x00063

Benso, S. (2000). *The face of things: A different side of ethics*. SUNY.

Bentham, J. (1988). *An introduction to the principles and morals of legislation*. Prometheus Books.

Birch, T. H. (1993). Moral considerability and universal consideration. *Environmental ethics, 15*, 313–332.

Calarco, M. (2008). *Zoographies: The question of the animal from Heidegger to Derrida*. Columbia University Press.

Chesterman, S. (2021). *We, the robots? Regulating artificial intelligence and the limits of the law*. Cambridge University Press.

Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. Palgrave Macmillian.

Deleuze, G. (1994). *Difference and repetition* (P. Patton, Trans.). Columbia University Press.

Derrida, J. (1978). *Writing and difference* (A. Bass, Trans.). University of Chicago Press.

Derrida, J. (2008). *The animal that therefore I am* (M. L. Mallet, Ed., and D. Wills, Trans.). Fordham University Press.

Descartes, R. (1988). *Selected philosophical writings* (J. Cottingham, R. Stoofhoff, & D. Murdoch, Trans.). Cambridge University Press.

Evans, E. P. (1906). *The criminal prosecution and capital punishment of animals*. William Heinemann.

Floridi. L. (2008). Information ethics, its nature and scope. In J. Van Den Hoven & J. Weckert (Eds.), *Information technology and moral philosophy* (pp. 40–65). Cambridge University Press.

Floridi, L. (2013). *The ethics of information*. Oxford University Press.

Fox, W. (1995). *Toward a transpersonal ecology: Developing new foundations for environmentalism*. SUNY Press.

Gellers, J. C. (2021). *Artificial intelligence, animal and environmental law*. Routledge.

Gunkel, D. J. (2012). *The machine question: Critical perspectives on AI, robots and ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Gunkel, D. J. (2020). *How to survive a robot invasion: Rights, responsibility and AI*. Routledge.

Hall, J. S. (2001, July 5). Ethics for machines. KurzweilAI.net. http://web.archive.org/web/20111031070519/http://www.kurzweilai.net/ethics-for-machines

Haraway, D. (1991). *Simians, cyborgs, and women: The reinvention of nature*. Routledge.

Heidegger, M. (1977). *The question concerning technology* (W. Lovitt, Trans.). Harper Torchbooks.

Hogan, K. (2017). Is the machine question the same question as the animal question? *Ethics and Information Technology, 19*(1), 29–38. https://doi.org/10.1007/s10676-017-9418-4

Kadri, S. (2007). *The trial: Four thousand years of courtroom drama*. Random House.

Kant, I. (1985). *Critique of practical reason* (L. W. Beck, Trans.). Macmillan Publishing Company.

Levinas, E. (1969). *Totality and infinity: An essay on exteriority* (A. Lingis, Trans.). Duquesne University.

Lingis, A. (1994). *The community of those who have nothing in common*. Indiana University Press.

Lippit, A. M. (2000). *Electric animal: Toward a rhetoric of wildlife*. University of Minnesota Press.

Lyotard, J. F. (1993). *The postmoden condition: A report on knowledge* (G. Bennington & B. Massumi, Trans.). University of Minnesota Press.

McGowan, T. (2014). Symptom. In R. Butler (Ed.), *The Žižek dictionary* (pp. 242–245). Acumen.

Moore, G. E. (2005). *Principia ethica*. Barnes & Noble Books.

Nadeau, J. E. (2006). Only androids can be ethical. In K. M. Ford, C. Glymour, & P. J. Hayes (Eds.), *Thinking about android epistemology* (pp. 241–248). IAAA Press.

Newton, I. (1972). *Philosophiae naturalis principia mathematica*. A. Koyre and I. B. Cohen (Eds.). Harvard University Press.

Nietzsche, F. (1966). *Beyond good and evil*. W. Kaufmann (Trans.). Vintage Books.

Philippopoulos-Mihalopoulos, A. (2017). Critical environmental law as method in the Anthropocene. In L. J. Kotzé (Ed.), *Reimagining environmental law and governance for the Anthropocene* (pp. 117–135). Hart.

Plato. (1977). *Protagoras*. W. R. M. Lamb (Trans.). Harvard University Press.

Plato. (1982). *Apology*. H. N. Fowler (Trans.). Harvard University Press.

Regan, T. (1999). Forward to P. Steeves *Animal others: On ethics, ontology, and animal life* (pp. xi–xiii). SUNY.

Sallis, J. (2010). Levinas and the elemental. In P. Atterton & M. Calarco (Ed.), *Radicalizing Levinas* (pp. 87–94). SUNY Press.

Sidgwick, H. (1981). *The methods of ethics*. Hackett Publishing Company.

Singer, P. (1973, April 5). Animal liberation. *The New York review of books, 20*(5). http://web.archive.org/web/20110109034541/http://www.nybooks.com/articles/archives/1973/apr/05/animal-liberation/

Singer, P. (1989). All animals are equal. In T. Regan & P. Singer (Eds.), *Animal rights and human obligations*, (pp. 148–162). Prentice-Hall.

Žižek, S. (1989/2008). *The sublime object of ideology*. Verso.

Žižek, S. (1999). *The ticklish subject*. Verso.

Žižek, S. (2000). *The fragile absolute*. Verso.

Human, Hybrid, or Machine?
Exploring the Trustworthiness of Voice-Based Assistants

Lisa Weidmüller[1]

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

Abstract

This study investigates how people assess the trustworthiness of perceptually hybrid communicative technologies such as voice-based assistants (VBAs). VBAs are often perceived as hybrids between human and machine, which challenges previously distinct definitions of human and machine trustworthiness. Thus, this study explores how the two trustworthiness models can be combined in a hybrid trustworthiness model, which model (human, hybrid, or machine) is most applicable to examine VBA trustworthiness, and whether this differs between respondents with different levels of prior experience with VBAs. Results from two surveys revealed that, overall, the human model exhibited the best model fit; however, the hybrid model also showed acceptable model fit as prior experience increased. Findings are discussed considering the ongoing discourse to establish adequate measures for HMC research.

Keywords: voice-based assistant, trustworthiness, trust, hybrid, scale, survey, prior experience

Introduction

In Human-Machine Communication (HMC) research, we are dealing with interactions between humans and perceptually *hybrid* communicative technologies such as voice-based assistants (VBAs). This means that humans who communicate and interact with these

technologies often do not perceive them solely as machines. Instead, because machines are increasingly built and programmed to imitate human capabilities and behavior "by exchanging messages with people or by performing a communicative task on their behalf" (Guzman, 2020, p. 37), they exhibit social cues that may prompt the attribution of human traits or social behavior to them (A. P. Edwards, 2018; Gambino et al., 2020; Garcia et al., 2018; Nass & Moon, 2000; Reeves & Nass, 1996). In fact, humans often perceive these technologies as "social things" (Guzman, 2015) or "personified things" (Etzrodt & Engesser, 2021), ascribing both human and machine characteristics to them.

Because of this perceptual hybridity, we need to adapt the way we investigate interactions with these communicative technologies and their effects (Chita-Tegmark et al., 2021; J. Edwards & Sanoubari, 2019). One context in which this is particularly relevant is the formation of trust toward communicative technologies. Trust has been identified as a crucial factor in the use and adoption of technology (e.g., Gao & Bai, 2014; C.-J. Lee et al., 2005; Priest, 2001). However, which trustworthiness characteristics form the basis for trust development depends on the perceived nature of the trustee, which is understood to differ substantially between human and machine trustees (Akter et al., 2011; Colquitt et al., 2007; Lankton et al., 2015; J. D. Lee & See, 2004). Therefore, communicative technologies like VBAs that blur perceptual categories, such as human and machine, are especially challenging.

VBAs, like Amazon's Alexa, Apple's Siri, or Google Assistant, represent a communicative technology that exhibits primary social cues due to its human-like conversational user interface (CUI) (Lombard & Xu, 2021). This triggers social reactions, interpersonal interaction patterns, and the attribution of (social) agency (Burgoon et al., 1999; Gambino et al., 2020; Lombard & Xu, 2021; McTear et al., 2016). Increasingly, VBAs also adapt a communicative role similar to that of a traditional news anchor. By selecting news content from external sources (e.g., media companies) and presenting the stories in their own voice and character in answer to a search query, VBAs take on the role of *artificial news anchors*. Their trustworthiness is therefore of particular importance because people may rely on the information they receive from VBAs to form opinions about the world surrounding them. As more people use VBAs to learn about the news (Ammari et al., 2019; Kinsella & Mutchler, 2020; Natale & Cooke, 2020; Newman, 2018) and VBA developer companies work to expand their respective VBAs' abilities to present news (Lyons, 2020; Porter, 2019), this use context, including VBAs' related trustworthiness, may deserve more attention than it currently receives (J. Edwards & Sanoubari, 2019). However, in this context, there currently exists no trustworthiness measure that adequately recognizes VBAs' perceptually hybrid nature as trustees and communicators.

Thus, the purpose of this study is to (1) empirically explore how the distinct human and machine trustworthiness models can be combined in a hybrid trustworthiness model for VBAs as artificial news anchors, (2) examine which model (human, hybrid, or machine) is most applicable to examine VBA trustworthiness in this context, and (3) determine whether this differs between respondents with different levels of prior experience with the VBA (i.e., no prior experience, indirect experience, or direct experience).

For these purposes, the paper, first, explicates what makes VBAs perceptually hybrid trustees and discusses what this means for assessing VBAs' trustworthiness. Second, a

hybrid model for VBA trustworthiness is exploratively developed by using data from two online surveys. Third, the competing human, machine, and hybrid models are empirically tested and compared to answer the following questions:

> **(RQ1)** Which trustworthiness model (human, hybrid, or machine) exhibits the best model fit and is thus most applicable to investigate VBA trustworthiness?

> **(RQ2)** How does people's level of prior experience with the VBA affect which trustworthiness model is most applicable?

Finally, implications of the findings for HMC research are discussed, and limitations, as well as directions for future research, are presented.

The Perceptual Hybridity of Voice-Based Assistants

VBAs can be distinguished from previous assistance applications by their sophisticated voice interface and dialogue system (Yang et al., 2019). Based on automatic speech recognition, analysis (natural language processing), synthesis (text-to-speech), and artificial intelligence (AI), VBAs can recognize and understand spoken instructions after receiving a wake word and can give meaningful answers or present information relevant to a query in a human-sounding voice (Deloitte, 2018; McTear et al., 2016). They are thereby able to mirror human interaction patterns, exhibit primary social cues, and simulate intentions through effective and meaningful behavior (Burgoon et al., 1999; Hearst, 2011; Lombard & Xu, 2021; Nass & Moon, 2000). This illusion of a human communicator is assisted by scripted small talk responses that are designed to convey a unique character or persona for each VBA (Natale, 2021), challenging the perception of VBAs as mere machines.

Studies have shown that the VBAs' human-like CUI activates scripts of interpersonal communication (Burgoon et al., 1999; McTear et al., 2016; Moon et al., 2016), prompting users to react socially and attribute human-like traits, like gender or personality, to the VBAs (Etzrodt & Engesser, 2021; Garcia et al., 2018; Guzman, 2019). Scholars have explained this behavior toward machines as a result of people's doubt (e.g., Reeves & Nass, 1996) about the machines' ontological classification. In essence, it is difficult for people to determine "who" or "what" a VBA is (Gunkel, 2020), or whether a VBA is a thing/object or person/subject (Etzrodt & Engesser, 2021). Most notably, people are in doubt about whether a VBA is a *human* or *machine* (Guzman, 2020). Recently, Etzrodt and Engesser (2021) explored the nature of this doubt, uncovering that, rather than assimilating VBAs into one of the schemes, people accommodated their schemes by classifying VBAs as "personified things." In other words, people ascribed both human and machine characteristics to VBAs, though not in the same amount (Etzrodt & Engesser, 2021). Similar results have been found for personified robots, which were attributed—to a certain degree—mental states, sociality, and even morality (Kahn, Jr. et al., 2011).

According to the Oxford English Dictionary (Oxford University Press, n.d.), something that is a mixture of (at least) two different elements is *hybrid*. Thus, VBAs can be understood as perceptual hybrids on the borderline of the human-machine distinction. They thereby

challenge existing models of trustworthiness and trust, which are suggested to differ substantially when dealing with a machine(-like) versus a human(-like) trustee (Lankton et al., 2015; J. D. Lee & See, 2004).

Trustworthiness and Trust Development

Trust is relevant in any situation where two or more distinct parties have to rely on one another to successfully complete a task or interaction involving uncertainty, unequally distributed knowledge, and/or the risk of negative consequences (Akter et al., 2011; J. D. Lee & See, 2004; Mayer et al., 1995; McKnight et al., 2011). Trust is therefore not limited to interpersonal exchanges but also includes human-machine interactions, and human-organization interactions (Schaefer et al., 2016). While the basic definition for trust is similar across these interactions, there are also differences, especially between human-human trust, and human-machine trust.

Trust can be defined, both in human-human interactions and human-machine (e.g., robot, automation, or agent) interactions, as an *attitude* of someone (hereafter, trustor) toward something or someone else (hereafter, trustee), coupled with the expectation that relying on the other party will prompt favorable outcomes (Blöbaum, 2016; Colquitt et al., 2007; J. D. Lee & See, 2004; Rousseau et al., 1998). Therefore, trust is not an action or behavior itself, but it can lead to behavioral intentions from which trust-related behavior (e.g., interaction, cooperation, or reliance) may result.

Over time, trust develops based on *trusting beliefs*, which are the result of the trustor's *experience from prior interactions* with a trustee and *immediately perceivable information* about the trustee's nature (J. D. Lee & See, 2004; Mayer et al., 1995; Rousseau et al., 1998). These trusting beliefs are often studied under the term *trustworthiness.* Trustworthiness can be defined as the trustor's attribution and evaluation of a trustee's abilities and characteristics that, the trustor believes, will lead to a beneficial outcome in a *specific context* (J. D. Lee & See, 2004; Mayer et al., 1995; McKnight et al., 2002).

The Role of the Trustee's Nature

The primary difference between human-human trust, and human-machine trust is the *nature of the trustees*. Unlike in human-human interactions, where both interaction partners belong to the same ontological category, in human-machine interactions, they do not (Guzman, 2020), making it more difficult for human trustors to determine the nature of their machine interaction partner. Because the nature of a trustee determines which of the trustee's characteristics are relevant to assess trustworthiness (Lankton et al., 2015; J. D. Lee & See, 2004), the perceptual hybridity of VBAs between human and machine challenges previously distinct definitions of human and machine trustworthiness.

A widely acknowledged definition of human-human trustworthiness—which guides the understanding of human-like trustworthiness in this study—is based on the following three dimensions: *integrity*, *competence,* and *benevolence* (Blöbaum, 2016; Mayer et al., 1995). To assess a human(-like) trustee as trustworthy thereby means that the trustee is perceived to adhere to certain moral and ethical values that are important to the trustor,

have the skills and characteristics to fulfill a certain task skillfully, and intent to do good to the trustor.

On the other hand, a widely acknowledged definition of trust in technology, or system-like trust—which guides the understanding of machine-like trustworthiness in this study—is based on the following three dimensions: *reliability*, *functionality,* and *helpfulness* (Chita-Tegmark et al., 2021; Lankton et al., 2015; McKnight et al., 109

). To assess a machine(-like) trustee as trustworthy thereby means that the trustee is perceived to consistently operate properly, have the functions or features needed to fulfill a specific task, and provide adequate help and assistance to the trustor.

These distinct definitions are based on the assumption that human(-like) trustees have the power to choose and make ethical decisions (i.e., intentions and volition), while machine(-like) trustees do not—apart from how they are preprogrammed (Lankton et al., 2015; J. D. Lee & See, 2004; Mayer et al., 1995). Thus, human trustees can choose to adhere to certain values or do good to the trustor, or they can choose not to. Machine trustees, on the other hand, cannot choose to operate properly or provide adequate assistance. Instead, machines need to have been correctly designed by human programmers to function without error and to provide the help needed by the user—without any intentionality on the part of the machine being involved. However, research shows that users can attribute intentionality to a technology itself (Guzman, 2019; Reeves & Nass, 1996; Sundar & Nass, 2000), especially when the technology exhibits sophisticated social cues that imply human-likeness (Gambino et al., 2020; Nass & Moon, 2000). Therefore, human and machine trustworthiness dimensions have to be combined to investigate the trustworthiness of such perceptually hybrid trustees (Chita-Tegmark et al., 2021; J. Edwards & Sanoubari, 2019).

The Role of Context

In addition to the differences between human and machine trustworthiness, which are based on the *nature of the trustee*, trustworthiness is *domain specific* (Mayer et al., 1995). That means that the goal or context of the interaction between the trustor and trustee determines which of the trustee's characteristics are relevant for the trustor's assessment of the trustee's trustworthiness (J. D. Lee & See, 2004).

In the context of presenting news, trustworthiness has been widely studied under the term *credibility* regarding the three domains of source, medium, and message. For investigating VBA trustworthiness, both the definitions of source and media credibility are relevant because the VBAs' human-like CUI has also been found to cause uncertainty about their role in the communication process. Some users, for instance, are inclined to identify VBAs as veritable sources rather than mediating channels of a message (Guzman, 2019). This is in line with the CASA (*Computers Are Social Actors*) paradigm, which purports that technologies that exhibit sufficient social cues—such as a human-sounding voice—can be perceived as the *source* of communication (Gambino et al., 2020; Lombard & Xu, 2021; Nass & Moon, 2000; Sundar & Nass, 2000).

Both the definitions of source and media credibility consistently include the dimension of expertise, defined as the *ability to know the truth*, and a dimension called trustworthiness, defined as the *motivation to tell the truth* (e.g., Hovland et al., 1953; Metzger et al.,

2003). Thus, both credibility dimensions emphasize the importance of *truthfulness*, which relates to the credibility assessment of the message. Since this study investigates how people assess the trustworthiness of VBAs in the context of presenting news, the human-like and machine-like trustworthiness dimensions need to be adapted to this specific context by incorporating the characteristic of truthfulness and the reference to the presented message.

VBAs as Hybrid Trustees

As aforementioned, research has shown that VBAs are perceptual hybrids. While undeniably technological devices, their human-like CUI causes uncertainty about whether VBAs are ontologically human or machine and, consequently, people attribute both human and machine characteristics to them (e.g., Etzrodt & Engesser, 2021). Additionally, this causes an overlap between the attributed role in the communication process as source or channel (e.g., Guzman, 2019). To incorporate this perceptual hybridity when examining VBA trustworthiness, this study explores how a hybrid trustworthiness model as a mix of human(-like) and machine(-like) trustworthiness attributes adapted to the news presentation context might look. Therefore, an exploratory factor analysis (EFA) will be conducted with items representing all six trustworthiness dimensions (human: integrity, competence, benevolence; machine: reliability, functionality, helpfulness). Theoretically, it would be possible that in a hybrid trustworthiness model, all six dimensions must be retained, or that two dimensions—a human and a machine trustworthiness dimension with the respective items—could evolve. However, it would also be possible that the characteristics from different dimensions of human and machine trustworthiness could mix and form hybrid dimensions of trustworthiness. The number of such hybrid dimensions is also open, although a three-dimensional structure would correspond to the established three-dimensional structure for both human and machine trustworthiness.

   After a hybrid model—of whatever form—is developed, the model fit of the competing models will be compared to investigate *which model is most applicable to investigate VBA trustworthiness (RQ1).*

The Moderating Role of Prior Experience

The literature suggests that prior experience plays an important role in establishing trust toward *any* trustee (J. D. Lee & See, 2004; Mayer et al., 1995; McKnight et al., 2002; Rousseau et al., 1998; Schaefer et al., 2016). Thus, we can distinguish between *initial trust* in an unfamiliar trustee (e.g., when the trustor first interacts with a trustee) and trust that is established based on prior interactions and experience with a trustee (Li et al., 2008; McKnight et al., 2002). Without having firsthand experience with a trustee, a trustor can still form an initial trustworthiness impression based on past experiences with similar trustees (Schaefer et al., 2016) or indirect experiences with the trustee through secondhand information from organizational and cultural contexts (J. D. Lee & See, 2004). Therefore, indirect experience with VBAs through advertisements, fiction, news, or gossip from others may also inform people's expectations of VBAs and thus their trustworthiness assessment.

   Previous research uncovered that people's expectations (i.e., mental models) of technologies that exhibit social cues change with growing experience (A. P. Edwards et al., 2019;

Gambino et al., 2020; Horstmann & Krämer, 2019; Lankton et al., 2014). Consequently, it is likely that people who have no prior experience with a VBA (i.e., those who can neither draw upon firsthand nor secondhand experience), those with indirect (secondhand) prior experience, and those with direct (firsthand) prior experience have different expectations of VBAs and thus assess their trustworthiness differently. For example, those with no prior experience or indirect experience possibly overestimate the human-likeness of VBAs and, thus, the human model could be most applicable. For the first group, this could simply be due to a novelty effect, for the latter it could be due to secondhand information from fiction or advertisements, where the human-likeness of such technologies is often emphasized. People with direct experience on the other hand, may have already adapted their mental models and, thus, the hybrid trustworthiness model might be most applicable. This will be examined with the following research question:

> **(RQ2)** How does people's level of prior experience with the VBA affect which trustworthiness model is most applicable?

Method

Sample

To investigate these research questions, an online survey was conducted in late 2018. Respondents were recruited among the students of a large German university via the university's email list. After eliminating incomplete cases (dropout rate: 2.7%) and cases of low quality according to the relative speed of completion and percentage of missing answers (Leiner, 2019), the final sample consisted of $N = 853$ students (response rate: 2.6%). On average, the respondents were 23 years old ($SD = 4.81$, age range = 17–50), and they were almost evenly split between men (52%) and women (48%). Most of the respondents were undergraduates (76%), and the vast majority already knew the VBA to which they were randomly assigned in the survey (Alexa: 96%, Google Assistant: 72%). However, for both VBAs, the students' knowledge primarily stemmed from indirect sources such as advertisements, other people, the media, or fiction. Only a few said they owned the assigned VBA themselves (Alexa: 7%, Google Assistant: 33%).

To validate the findings from this pilot study, the survey was additionally conducted with a slightly older sample. Therefore, staff members of the same university were recruited via the university's staff email list (dropout rate: 7.4%, response rate: 6.2%). On average, staff respondents ($N = 435$, 53% male) were 10 years older (mean age = 33, $SD = 10.35$, age range = 18–65) than respondents in the student sample. Furthermore, 69% of staff respondents were graduates compared to only 24% in the student sample. Both samples therefore represent two slightly different educational phases: those still in higher education (students) versus those who have finished higher education (staff). Regarding prior experience with the VBAs, the staff sample largely resembled the student sample, but fewer knew their assigned VBA (Alexa: 93%, Google Assistant: 65%) or owned the VBA (Alexa: 6%, Google Assistant: 24%). Both samples will be used to validate the exploratively derived hybrid trustworthiness model and account for possible cohort effects.

Procedure

Because the level of VBA adoption in Germany was still low at the time of the survey, with only 26% of German internet users using VBAs in general (Taş et al., 2019) and only 5% using smart speakers (Newman et al., 2018), a *demonstrational survey design* was chosen. After asking about respondents' prior experience with several VBAs, they were shown pre-recorded videos of either the smart speaker variant of Google Assistant (Google Home) or Alexa (Amazon Echo) within the online survey. These two VBAs were selected due to their leading market position (Kinsella & Mutchler, 2020; Statista, 2021); however, the random assignment to either one of the VBAs is not essential for the paper at hand. The videos were between 11 and 19 seconds long and were presented in a way that simulated inter-actions with the respective VBA in a news use scenario. All respondents successively acti-vated three predefined, news-related questions by clicking a button. Then, they received the VBAs' answers in the form of the prerecorded videos, which are available in this study's OSF repository (including English transcripts).

Though this presentation mode limits the actual conversational nature of interacting with VBAs, it was chosen for three reasons. First, while both VBAs were already available on the German market, making direct or indirect experience with them possible, adoption in Germany was still low. Thus, the prerecorded videos allowed respondents to get a good impression of how VBAs present news.[1] Second, VBA functionality for presenting news was still in its infancy in Germany at the time of the survey. Thus, the prerecorded videos made it possible to simulate a functionality, which was already a reality in English-speaking coun-tries, while still using VBAs that were available on the German market. Third, VBAs present different content depending on which VBA is used, the time and date, and how a search query is formulated. The use of predefined questions and prerecorded videos was therefore necessary to ensure that all respondents received the same content, no matter which VBA they were assigned or when they participated in the survey. The content was selected from real news of quality German media, which the VBAs were manipulated to present verbally, without citing the source, by using the IFTT ("If This Then That")-App and then recording the answers. For the topic, the implementation of the General Data Protection Regulation (GDPR) was chosen due to its societal relevance in Germany at the time of the study. Addi-tionally, the topic was associated with a certain level of risk and uncertainty because of possible (negative) consequences if the regulations were not implemented correctly. As risk contributes to the relevance of trust(worthiness) (e.g., J. D. Lee & See, 2004; McKnight et al., 2002), the topic provided respondents with the necessary incentive to assess trustwor-thiness, which respondents were asked to do after they saw the videos.

Measures

VBA Trustworthiness

VBA trustworthiness was measured by asking respondents to indicate how much they agreed with 16 items on a seven-point Likert scale (1 = "Do not agree at all" to 7 =

---

1. Four additional prerecorded videos that contained the VBAs' answers to predefined "personal" questions (e.g., "How are you?") were included in the beginning of the survey (after asking about prior experience) so that respondents could get a good impression of the respective VBA's persona, as well.

"Agree completely"). Additionally, they were able to choose a no response option: "I cannot assess that."

Items were developed to represent the definitions of the three human-like trustworthiness dimensions: integrity, competence, and benevolence (Mayer et al., 1995), and the three machine-like trustworthiness dimensions: functionality, reliability, and helpfulness (Lankton et al., 2015; McKnight et al., 2011). Most of the items are based on existing measures used to investigate trust in technology (Fink, 2014; Lankton et al., 2015), e-commerce (Koh & Sundar, 2010; McKnight et al., 2002; Wang & Benbasat, 2016), anthropomorphic agents (Burgoon et al., 1999), human news anchors, journalists (Newhagen & Nass, 1989), and human endorsers (Ohanian, 1990). However, the wording of the items was specifically adapted to the trustee (VBAs) and the context of information presentation.

Human-Like Trustworthiness Items.   For integrity, respondents were asked to assess whether the VBA was *trustworthy* and *credible* in general and whether the VBA *gives truthful information*. To be influential within the news presentation context (competence), the VBA had to be assessed as being *competent* and *qualified to provide information*, as well as *reliable*. Benevolence of the VBA was operationalized as *acting in the trustor's interest*, being *interested in the trustor's well-being*, and *providing help if needed*.

Machine-Like Trustworthiness Items.   When developing the machine-like trustworthiness items, previous scales were consulted (Lankton et al., 2015; McKnight et al., 2011; Ullman & Malle, 2018). However, the items had to be adapted to the specific trustee and trusting context. Specifically, to measure whether respondents believed that the VBA consistently operates properly, (reliability) items included were *is reliable*, *gives security in information search,* and *respects people's privacy*, thereby incorporating common user (and designer) concerns regarding VBAs (Clark et al., 2019; Easwara Moorthy & Vu, 2015; Kinsella & Mutchler, 2020; Lei et al., 2018; Olson & Kemery, 2019). Items for the VBAs' functionality to provide information were based on aspects of understandability (*understandable and fluent presentation*) and ease of use (*intelligent* and *well-programmed*). Helpfulness was defined as providing help and advice necessary to fulfill a task, which in this context means selecting and presenting relevant news in response to a query. Therefore, respondents had to indicate whether they believed the VBA *is useful*, *gives relevant answers*, and *provides help if needed*.

An overview of the items' wordings and the underlying definitions is given in Table 1.[2] All of the items will serve as the basis for the exploratory development of the hybrid trustworthiness model.

---

2. Conceptually, two items belong to both a human-like and a machine-like trustworthiness dimension. First, to be *reliable* was developed as a competence characteristic within the domain of news presentation (human-like: competence), and to reflect the characteristic to consistently operate properly (machine-like: reliability). Second, the definitions of benevolence (human-like) and helpfulness (machine-like) share the similar notion of *providing help if needed*.

TABLE 1	Definition and Operationalization of Human- and Machine-Like Trustworthiness Dimensions Adapted to the Context of News Presentation				
Human-Like Trustworthiness			**Machine-Like Trustworthiness**		
Dimensions		**Items**	**Dimensions**		**Items**
Integrity	The belief that a trustee adheres to a set of principles that the trustor finds acceptable *(domain-specific)*	• can be trusted • is credible • gives truthful information	**Reliability**	The belief that the specific technology will consistently operate properly	• is reliable • provides security in information search • respects people's privacy
Competence	The belief that the trustee has the skills, competencies, and characteristics needed to have influence within a *specific domain*	• is reliable • is competent in providing information • is qualified to provide information	**Functionality**	The belief that the specific technology has the capability, functionality, or features to do for one what one needs to be done *(task-specific)*	• answers understandably & fluently • is well-programmed • is intelligent
Benevolence	The belief that the trustee will want to do good to the trustor apart from an egocentric motive	• acts in my best interest • is interested in my well-being • would do its best to help me if I needed help	**Helpfulness**	The belief that the technology provides adequate, effective, and responsive help or advice necessary to successfully complete a *specific task*	• is useful • gives relevant answers • would do its best to help me if I needed help
Notes. Based on Burgoon et al., 1999; Fink, 2014; Koh & Sundar, 2010; Lankton et al., 2015; Mayer et al., 1995; McKnight et al., 2002; McKnight et al., 2011; Newhagen & Nass, 1989; Ohanian, 1990; Wang & Benbasat, 2016. Items were translated from German original (see supplemental material).					

Prior Experience

Prior experience with the respective VBA was measured by asking whether respondents knew the assigned VBA before the study (yes/no) and from where they had learned about the VBA (e.g., advertisements, other people, the media, fiction, or ownership). Based on these measures, respondents of both samples were grouped according to their prior experience with the assigned VBA: (1) those who did not know the respective VBA before the study were categorized as having *no prior experience,* (2) those who knew the VBA before the study but only through indirect contact (e.g., advertisements, other people, the media,

or fiction) were categorized as having *indirect experience,* and (3) those who said they owned the VBA were categorized as having *direct experience.* Among student respondents, 16% had no prior experience with their assigned VBA, 64% had indirect experience, and 20% said they had direct experience. This was similar in the staff sample, where 21% of respondents had no prior experience, 64% had indirect experience, and 15% said they had direct experience. This means that in both samples, most respondents had no or indirect prior experience with the VBA they had to assess in the survey. This was anticipated due to the low level of adoption of VBAs in Germany at the time of the survey (Newman et al., 2018; Taş et al., 2019), which is why the demonstrational survey design was employed. For these respondents, the assessment likely represents a first impression, and findings must be interpreted accordingly.

Results

This section is divided into four stages of inquiry. First, the percentages of the "I cannot assess that" responses to all individual trustworthiness items were examined and interpreted as a first indication of respondents' ability to apply the proposed items to VBAs in the context of news presentation. Second, to explore how a hybrid trustworthiness model for VBAs might look, an exploratory factor analysis (EFA) was conducted in the student sample with all 16 trustworthiness items from both the human-like and machine-like trustworthiness dimensions. The resulting model was then submitted to a confirmatory factor analysis (CFA) in the staff sample, to cross-validate it with this different cohort. Third, the applicability of the competing models—human, hybrid, or machine **(RQ1)**—was compared by conducting CFAs and contrasting the resulting model fit statistics. This procedure was applied in both samples to account for possible cohort effects. Last, the model comparison was repeated for different subgroups of respondents, which were specified according to the respondents' prior experience (i.e., none, indirect, or direct) with the VBA **(RQ2)**. Supplemental materials are available in this study's OSF repository.

Applicability of Individual Trustworthiness Items

As is apparent from Figure 1, most respondents from both samples were able to apply most of the items to VBAs. Student respondents had difficulties applying the human-like trustworthiness item *gives me truthful information*, as well as the two machine-like items *respects people's privacy* and is *well-programmed*. All of these may be caused by a lack of prior experience with the VBA in the context of news presentation, or a lack of contextual information available from the simulated interaction within the survey (Chita-Tegmark et al., 2021). Staff respondents had difficulties applying the same items but also struggled to assess the human-like item *acts in my best interest*, as well as both items that are conceptually part of both the human- and machine-like trustworthiness model: *is reliable* and *would do its best to help me if I needed help*. While the difficulty of assessing reliability may again result from a lack of experience or lack of contextual information, staff respondents' difficulties with the other two items may have been caused by a lack of perceived agency (Chita-Tegmark et al., 2021). Staff respondents may simply not perceive VBAs as having the ability to act, and, thus, they might not think that VBAs can help them.

FIGURE 1   Percentage of "I cannot assess that" Responses to VBA Trustworthiness Items
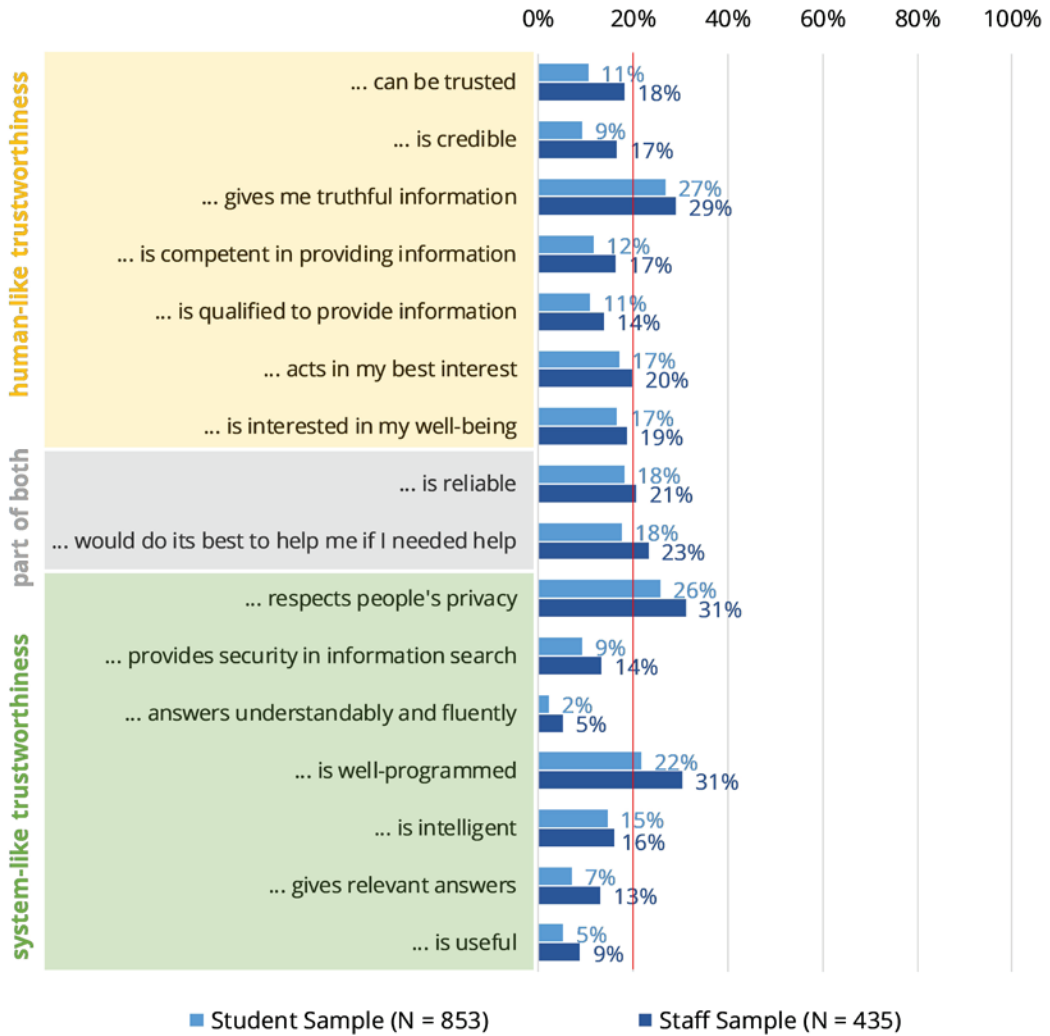
Note. Those items with percentages over 20% (red line) are discussed in detail.

In summary, more machine-like than human-like trustworthiness items caused assessment problems in both samples, which may be a first hint that they are not applicable to VBAs in the context of news presentation. Additional statistics (e.g., means, standard deviations, skewness, and kurtosis) are provided in the supplemental material.

Exploring the Hybrid Trustworthiness Model

To explore how a hybrid trustworthiness model including both human-like and machine-like items might look for VBA trustworthiness, an EFA was conducted with the student sample. The resulting model was then subjected to CFA in the staff sample to cross-validate it.

Student Sample: Exploratory Factor Analysis

First, the correlation matrix (supplemental material) for all items was examined. Thus, two items had to be excluded due to their low correlation (< 0.3) with more than one third of the other items, which suggests that they "do not 'fit' with the pool of items" (Field, 2018, p. 806). This concerned two machine-like items: the reliability item *respects people's privacy*, and the functionality item *answers understandably and fluently*. Both were excluded from the following EFA. Bartlett's test of sphericity ($\chi^2(91) = 2967.04$, p < .001), as well as the Kaiser-Meyer-Olkin measure (KMO = 0.935), confirmed the sampling adequacy for the analysis (Field, 2018). Because the literature suggests that the dimensions of trustworthiness correlate and interact with one another (Mayer et al., 1995), oblique rotation (promax, kappa = 4) was chosen to form factors. Items with communalities (after extraction) lower than 0.35 were excluded because low communalities suggest that the extracted factors only explain a small portion of variance in these items, sharing only a low level of common variance with the rest of the items (Field, 2018). This applied to the two machine-like functionality items *is well-programmed* and *is intelligent*.

The subsequent EFA with the remaining 12 items resulted in a two-factor solution (Table 2), according to Kaiser's criterion of eigenvalues over 1. Because machine-like reliability and helpfulness items, as well as human-like integrity and competence items, were clustered together on the first factor, this factor can be interpreted as *hybrid trustworthiness*, a true mix of both human and machine trustworthiness attributes. The second factor contained the three human-like *benevolence* items. Thus, for the student respondents, benevolence seems to stand out from the rest of the trustworthiness items when assessing VBAs, which may suggest that this dimension is inadequate for the trustworthiness assessment of these technologies. However, correlation between the two factors is high, with $r = .65$. Together, the two factors explained 62.71% of the variance.

Because of the established three-dimensional structure of previous trustworthiness models, the EFA was conducted again, this time forcing three factors to be extracted. As a result, the third extracted factor did not meet Kaiser's criterion, with an eigenvalue of 0.79. This eigenvalue is acceptable according to Jolliffe's more liberal criterion (Field, 2018). Therefore, the three-factor solution was maintained due to its agreement with previous trustworthiness definitions. Based on the way items were clustered together (Table 3), the three factors were interpreted as (1) *hybrid competence*, (2) *hybrid integrity*, and (3) *benevolence*—thus resembling dimensions found mostly in human trust literature, even if attributes from both human and machine trustworthiness were mixed. Together, the three factors explained 69.31% of the variance. Additionally, this solution improved the average communality after extraction. This factor solution is retained as an alternative hybrid trustworthiness model.

In summary, the EFAs in the student sample produced two alternative hybrid models for VBA trustworthiness.

	TABLE 2 Hybrid Trustworthiness Model: 2-Factor Solution From EFA With the Student Sample		
	Factor Loadings		
	Hybrid Trustworthiness	**Benevolence**	**Communalities**
... is credible	**0.86**	0.00	0.75
... is competent in providing information	**0.84**	-0.07	0.63
... gives me truthful information	**0.77**	-0.07	0.53
... is qualified to provide information	**0.73**	0.07	0.61
... is reliable	**0.71**	0.02	0.53
... provides security in information search	**0.70**	0.03	0.52
... can be trusted	**0.67**	0.13	0.59
... is useful	**0.67**	0.01	0.45
... gives relevant answers	**0.62**	0.05	0.42
... is interested in my well-being	-0.10	**0.76**	0.49
... would do its best to help me if I needed help	0.03	**0.73**	0.57
... acts in my best interest	0.14	**0.63**	0.54
Eigenvalues	6.31	1.12	6.63
% of variance	52.60	10.11	
Cronbach's α	0.91	0.78	

Notes. n = 443; KMO = .927; R^2 = 62.71%
Extraction Method: Principal Axis Factoring; Rotation Method: Promax with Kaiser Normalization
Factor loadings > 0,40 appear in bold

**TABLE 3   Hybrid Trustworthiness Model:
3-Factor Solution From EFA With the Student Sample**

	Factor Loadings			
	Hybrid Competence	**Hybrid Integrity**	**Benevolence**	**Communalities**
… is competent in providing information	**0.85**	0.06	-0.08	0.72
… is reliable	**0.80**	−0.04	0.02	0.62
… gives relevant answers	**0.55**	0.10	0.05	0.45
… is useful	**0.55**	0.15	0.01	0.47
… is qualified to provide information	**0.42**	0.35	0.07	0.60
… can be trusted	-0.07	**0.81**	0.12	0.69
… is credible	0.13	**0.81**	-0.02	0.82
… gives me truthful information	0.10	**0.74**	-0.09	0.59
… provides security in information search	0.27	**0.47**	0.03	0.52
… would do its best to help me if I needed help	0.21	−0.19	**0.76**	0.62
… is interested in my well-being	−0.17	0.09	**0.74**	0.49
… acts in my best interest	0.03	0.15	**0.61**	0.53
Eigenvalues	6.31	1.12	0.79	7.12
% of variance	52.60	10.11	6.61	
Cronbach's α	0.84	0.87	0.78	

Notes. n = 443; KMO = .927; R^2 = 69.31%
Extraction Method: Principal Axis Factoring; Rotation Method: Promax with Kaiser Normalization
Factor loadings > 0,40 appear in bold

Staff Sample: Confirmatory Factor Analysis

To cross-validate the two alternative hybrid trustworthiness models with the staff sample, confirmatory factor analyses (CFAs) were conducted using R (packages lavaan and semPlot). Both models were specified according to the findings from the EFAs in the student sample. The CFAs were interpreted by looking at a combination of recommended model fit statistics such as the comparative fit index (CFI > .95 for good fit), Tucker Lewis index (TLI > .95 for good fit), root mean square error of approximation (RMSEA < .08 for adequate fit), and standardized root mean square of residuals (SRMR < .06 for good fit) (Hu & Bentler, 1999; Kline, 2011; Pertegal et al., 2019; Weiber & Mühlhaus, 2014).

Based on the recommended thresholds for these fit statistics, the results of the two CFAs show that only the three-factor hybrid model exhibited an acceptable model fit in the staff sample, while the two-factor hybrid model did not meet the requirements for fit statistics (Table 4).[3] Thus, the two-factor hybrid model was rejected and only the three-factor hybrid model will be considered as the *hybrid model* in the following analyses.

TABLE 4 Hybrid Trustworthiness Model: Summary of Model Fit Statistics From CFAs With the Staff Sample		
	Staff Sample ($N = 435$)	
	hybrid 2 factors	**hybrid 3 factors**
N	207	207
χ^2	221.00	102.99
df	53	51
GFI	0.82	0.93
CFI	0.88	0.96
TLI	0.85	0.95
RMSEA	0.12	0.07
RMSEA LL 90% CI	0.11	0.05
RMSEA UL 90% CI	0.14	0.09
SRMR	0.06	0.05
Notes. Acceptable fit statistics are highlighted in light green. Not acceptable fit statistics appear in red. GFI – goodness of fit index >.90; CFI – comparative fit index >.95; TLI – Tucker-Lewis index >.95; RMSEA – root mean square error of approximation <.08; SRMR – standardized root mean square residual <.06		

---

3. The resulting standardized factor loadings from the CFAs are available in this study's OSF repository.

Applicability of Different Trustworthiness Models: Human, Hybrid, or Machine?

To compare the human(-like), machine(-like), and hybrid models of VBA trustworthiness, additional CFAs were conducted in both samples to validate and compare the findings of this pilot study between two slightly different cohorts. While the exclusively human and exclusively machine models were specified based on the respective theoretical trustworthiness definitions, the hybrid model was specified according to the three-factor model that was exploratively developed with the student sample and validated with the staff sample. The applicability of the competing models was evaluated and contrasted by looking at the same combination of recommended model fit statistics as before. Thus, a model is interpreted as *applicable* if the results of the CFA indicate that its model fit statistics are above (goodness of fit) or below (badness of fit) the recommended thresholds. The model that exhibits the best model fit is interpreted as *most applicable* (i.e., to work best).

In both samples, the human-like, as well as the hybrid trustworthiness model, demonstrated an acceptable model fit. Meanwhile, in both samples, the machine-like model did not meet the requirements for fit statistics. Comparing the human-like and the hybrid trustworthiness model showed that the human-like model exhibited an overall better model fit according to conventional fit statistics. These results were uncovered using the student sample (Table 5), and they were validated using the staff sample (Table 6). Thus, the findings apply for two different cohorts. This suggests that a human-like model is indeed suitable for investigating VBAs' trustworthiness in the context of news presentation *(RQ1)*.

TABLE 5 Model Comparison (RQ1): Summary of Model Fit Statistics From CFAs With the Student Sample		
Student Sample (*N* = 853)		
human-like	**machine-like**	**hybrid**
N	459	413	443
χ^2	86.16	101.70	169.16
df	24	24	51
GFI	0.96	0.94	0.94
CFI	0.97	0.93	0.96
TLI	0.96	0.89	0.95
RMSEA	0.08	0.09	0.07
RMSEA LL 90% CI	0.06	0.07	0.06
RMSEA UL 90% CI	0.09	0.11	0.08
SRMR	0.04	0.05	0.04

Notes. Acceptable fit statistics are highlighted in light green. Best fit statistics are highlighted in dark green. Not acceptable fit statistics appear in red.
GFI – goodness of fit index >.90; CFI – comparative fit index >.95; TLI – Tucker-Lewis index >.95; RMSEA – root mean square error of approximation <.08; SRMR – standardized root mean square residual <.06

TABLE 6 Model Comparison (RQ1): Summary of Model Fit Statistics From CFAs With the Staff Sample			
	Staff Sample ($N = 435$)		
	human-like	**machine-like**	**hybrid**
N	218	185	207
χ^2	38.53	67.07	102.99
df	24	24	51
GFI	0.96	0.92	0.93
CFI	0.99	0.90	0.96
TLI	0.98	0.85	0.95
RMSEA	0.05	0.10	0.07
RMSEA LL 90% CI	0.02	0.07	0.05
RMSEA UL 90% CI	0.08	0.13	0.09
SRMR	0.03	0.06	0.05

Notes. Acceptable fit statistics are highlighted in light green. Best fit statistics are highlighted in dark green. Not acceptable fit statistics appear in red.
GFI – goodness of fit index >.90; CFI – comparative fit index >.95; TLI – Tucker-Lewis index >.95; RMSEA – root mean square error of approximation <.08; SRMR – standardized root mean square residual <.06

The Moderating Role of Prior Experience

Finally, the competing models were compared regarding their fit according to the respondents' prior experience with the VBA. Model fit statistics for all comparisons described in this section are provided in the study's OSF repository. An overview of the applicability of all models according to the level of respondents' prior experience with the VBA is displayed in Table 7.

TABLE 7 Overview of Model Fit From CFAs According to Respondents' Prior Experience (RQ2)				
Sample	**Prior Experience**	**human-like**	**machine-like**	**hybrid**
Student	None ($n = 136$)			x
	Indirect ($n = 546$)			
	Direct ($n = 171$)			
Staff	None ($n = 90$)		x	x
	Indirect ($n = 280$)		x	
	Direct ($n = 65$)	x	x	x

Notes. Models with acceptable model fit are highlighted in light green. The model exhibiting the best model fit for each subgroup is highlighted in dark green. Not acceptable models according to fit statistics are marked by a red x.

For students who had no prior experience with the VBA before the study, the machine-like model was most applicable (i.e., exhibited the overall best model fit), while the human-like model also exhibited an acceptable model fit. For those with indirect experience, the human-like model was most applicable, but both the machine-like and the hybrid model showed an acceptable fit as well. For those who owned the VBA themselves and thus have direct experience with the VBA, the human-like trustworthiness model still exhibited the best model fit but slightly less than for those with indirect experience. The model fit of the machine-like and hybrid model did not change and was also acceptable for this subgroup.

In the staff sample, the human-like model worked best for those who did not know the VBA before the study, as well as those with indirect experience. For the subgroup with indirect experience, the hybrid model also showed an acceptable fit. However, no model exhibited an acceptable model fit for those who had direct experience with the VBA, which may be due to the small size of this subgroup in the staff sample ($n = 65$).

Nonetheless, these results indicate that experience—even indirect experience—affects which models are suitable to assess communicative technologies (**RQ2**), as has been proposed by other scholars (e.g., Gambino et al., 2020; Horstmann & Krämer, 2019). Because this study was conducted in an early stage of VBA adoption in Germany, even those with direct experience might be relatively inexperienced. Over time, they may adapt their models further, and the hybrid model may be more applicable still.

Discussion of Results and Contributions

This study is a pilot study that takes a first step toward understanding how we can adequately investigate trustworthiness in emerging technologies that are perceptually hybrid and take on roles that were previously inherent to humans (e.g., communicator, news anchor). Due to their human-like CUI, VBAs are often attributed a mixture of human and machine characteristics, making them perceptual hybrids on the borderline between human(-like) and machine(-like). The paper discussed the implications of this perceptual hybridity of VBAs for the assessment of trustworthiness. As a result, both human-like and machine-like trustworthiness dimensions were considered to explore how a hybrid trustworthiness model could look.

In this explorative study, two well-interpretable hybrid trustworthiness models as a mix of both human-like and machine-like characteristics were discovered, one of which—the hybrid model with three factors—showed an acceptable model fit. However, when comparing model fit for the human, hybrid, and machine models, the *human* model exhibited the overall best model fit in both samples, which speaks in favor of the validity of this finding across different cohorts. This indicates that scales developed for human-human interactions might be, after all, applicable to VBAs.

The study also explored whether the applicability of the models differed between respondents with different levels of prior experience with the VBA (no prior experience, indirect experience, or direct experience). For those with no prior experience, this differed substantially between the two samples: While for students, the *machine* model exhibited the best model fit, for staff respondents, only the *human* model had an acceptable model fit. It seems that when people encounter VBAs for the first time (in an online survey, no

less), which models are applicable, and thus how respondents interpret the available cues, depends largely on their predispositions.

When respondents of both samples had at least indirect experiences with the VBA, the human model exhibited the best model fit. Whether this is because the available second-hand information (e.g., advertisements or friends who just bought the VBA) emphasizes the human-likeness of VBAs and "triggers" this model is beyond the scope of this paper. However, the hybrid model also exhibited acceptable model fit at this level of prior experience, which may indicate that even secondhand information affects how hybrid communicators are assessed.

For students with direct experience with the VBA, the applicability of the human model decreased while that of the machine and hybrid model remained the same, thereby narrowing the fit gap between the competing models. Thus, it seems to depend on people's level of prior experience as to which characteristics they apply to assess trustworthiness when interacting with perceptually hybrid communicative technologies such as VBAs. Moreover, with increasing experience of the respondents, the applicability of the different models changed and started to match the perceptual hybridity of VBAs found in previous studies. These findings support the assumption that, as these hybrid technologies are increasingly embedded in our daily lives and routines, people's interaction scripts and mental models will evolve (e.g., Gambino et al., 2020; Horstmann & Krämer, 2019).

By testing alternative hybrid and non-hybrid models for VBA trustworthiness, this pilot study contributes to the ongoing discourse about adequate instruments for investigating communicative technologies. As is increasingly apparent, neither scales from social psychology developed for human-human interactions, nor those with a focus on usability or functionality, adequately recognize the nature of communicative technologies such as VBAs. Instead, they either over- or underestimate VBAs' perceptual human-likeness. While exploratory, this study demonstrated how the two previously distinct human and machine models of trustworthiness mix when the investigated trustee's nature blurs the distinction these models are based on (human-like or machine-like). Thus, this study provides further indications that previous scales may need to be adapted and new measures developed due to this perceptual hybridity. Future studies should apply a qualitative survey approach to explore whether there are additional, genuinely hybrid trustworthiness characteristics currently missing from the hybrid model.

Limitations and Directions for Future Research

The reported findings must be interpreted considering some limitations of the study. To achieve high internal validity for this pilot study and a large sample size, the data were collected via online surveys. Therefore, interactions with the VBAs had to be simulated by presenting prerecorded videos. It is possible that respondents may assess VBAs differently when directly interacting with them. Furthermore, this study was part of a larger project. In the beginning of both surveys, four additional videos were presented that contained the VBAs' answers to "personal" questions. These videos may have affected respondents' perceptions of the VBAs as more human-like, thus also possibly affecting the model fit of the human-like trustworthiness model. Hence, the hybrid model may have exhibited the best model fit if these videos had not been part of the survey. The connection between perception

and assessment should be further investigated, for example, by exploring whether people apply different trustworthiness models according to how they ontologically classify VBAs (i.e., human, machine, or something in between) and how they perceive the VBAs' role in the communication process (i.e., source or channel).

This study found indications for the widely proposed adaptation of mental models for communicative technologies with growing experience. However, both samples were very similar and mostly knew the VBAs from indirect contact, the results may be most conclusive for initial trustworthiness assessments. However, because the study used real VBAs that are available on the market, two out of three respondents had at least indirect experience with the assigned VBA, and one in five student (and one in seven staff) respondents had direct experience with the assigned VBA, which is not the case for most laboratory experiments using fictious assistants or chatbots. Future studies should investigate whether the findings also apply to a more experienced population. Thus, environmental factors, such as perceived competence of the programmers or developer companies of a VBA, should be included as well to explore how these factors interact with the trustworthiness assessment of the VBA itself and how this affects the development of trust. Furthermore, VBA assessment should be tested with older participants and participants with a more diverse educational background.

In terms of context, different levels of perceived risks and expectations may be involved when VBAs present different kinds of information. Thus, future studies should investigate the applicability of different trustworthiness models for different types of information (e.g., news or service). Additionally, a media trustworthiness model should also be considered. A comparison of different versions of the same VBA (e.g., text-based, voice-based, and voice-based plus display) might also yield important insights into how the trustworthiness assessment of VBAs in the context of news presentation is affected by presentation mode.

Conclusion

This study explored how the perceptual hybridity of communicative technologies entails the need for adequate hybrid measurement models. This was tested for a construct that emphasizes distinguishing between humans and machines but is nonetheless important for both: trustworthiness. This pilot study contributes to that by examining whether previously distinct models of human and machine trustworthiness can be applied to VBAs, or whether their combination in a hybrid trustworthiness model better captures the perceptual hybridity of VBAs. By examining the applicability of the competing human, machine, and hybrid trustworthiness models in the context of news presentation, the study provides insights into people's mental models when assessing these technologies. While overall the human model had the best model fit in this study, the findings also supported a hybrid trustworthiness model. Additionally, model applicability was found to change with different levels of prior experience. With no prior experience, either the human or machine model worked best, depending on which cohort the respondents belonged to. When respondents had at least some (either indirect or direct) experience with the technology, the human, machine, and hybrid models were simultaneously applicable, suggesting that respondents had already begun to adapt their mental models. Thus, this paper argues that there is a need to develop hybrid measures that adequately recognize the hybrid nature of communicative technologies such as VBAs. Future research is needed to explore if there are additional, genuinely

hybrid trustworthiness characteristics, and whether they supplement or replace human and machine trustworthiness characteristics.

Author Biography

Lisa Weidmüller is a PhD student and research associate at the Institute of Media and Communication at TU Dresden. Her research focuses on interaction with and through emerging digital technologies and information intermediaries, especially regarding acceptance and development of trust.

ⓘD https://orcid.org/0000-0003-1680-9122

Center for Open Science

References

Akter, S., D'Ambra, J., & Ray, P. (2011). Trustworthiness in mHealth information services: An assessment of a hierarchical model with mediating and moderating effects using partial least squares (PLS). *Journal of the American Society for Information Science and Technology*, *62*(1), 100–116. https://doi.org/10.1002/asi.21442

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), 1–28. https://doi.org/10.1145/3311956

Blöbaum, B. (2016). Key factors in the process of trust. On the analysis of trust under digital conditions. In B. Blöbaum (Ed.), *Trust and communication in a digitized world: Models and concepts of trust research* (pp. 3–25). Springer International Publishing. https://www.springer.com/de/book/9783319280578

Burgoon, J. K., Bonito, J. A., Bengtsson, B., Ramirez, A., Dunbar, N. E., & Miczo, N. (1999). Testing the interactivity model: Communication processes, partner assessments, and the quality of collaborative work. *Journal of Management Information Systems*, *16*(3), 33–56. https://doi.org/10.1080/07421222.1999.11518255

Chita-Tegmark, M., Law, T., Rabb, N., & Scheutz, M. (2021). Can you trust your trust measure? *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, 92–100. https://doi.org/10.1145/3434073.3444677

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., Aylett, M., Cabral, J., Munteanu, C., Edwards, J., & Cowan, B. R. (2019). The state of speech in HCI: Trends, themes, and challenges. *Interacting with Computers*, *31*(4), 349–371. https://doi.org/10.1093/iwc/iwz016

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

Deloitte. (2018). *Beyond touch: Voice-commerce 2030*. https://web.archive.org/web/20211228170629/https://www.deloitte-mail.de/custloads/141631293/Deloitte-Beyond-Touch-Studie.pdf

Easwara Moorthy, A., & Vu, K.-P. L. (2015). Privacy concerns for use of voice activated personal assistant in the public space. *International Journal of Human-Computer Interaction*, *31*(4), 307–335. https://doi.org/10.1080/10447318.2014.986642

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

Edwards, A. P., 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, J., & Sanoubari, E. (2019). A need for trust in conversational interface research. *Proceedings of the 1st International Conference on Conversational User Interfaces*, 1–3. https://doi.org/10.1145/3342775.3342809

Etzrodt, K., & Engesser, S. (2021). Voice-based agents as personified things: Assimilation and accommodation as equilibration of doubt. *Human-Machine Communication*, *2*, 57–79. https://doi.org/10.30658/hmc.2.3

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). SAGE Publications.

Fink, D. (2014). *Vertrauen in autonome Technik [Trust in autonomous technology]* [TU Dortmund]. http://dx.doi.org/10.17877/DE290R-15509

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

Gao, L., & Bai, X. (2014). A unified perspective on the factors influencing consumer acceptance of internet of things technology. *Asia Pacific Journal of Marketing and Logistics*, *26*(2), 211–231. https://doi.org/10.1108/APJML-06-2013-0061

Garcia, M. P., Lopez, S. S., & Donis, H. (2018, July 1). Voice activated virtual assistants personality perceptions and desires: Comparing personality evaluation frameworks. *Proceedings of British HCI 2018*. 32nd International BCS Human Computer Interaction Conference, Belfast, UK. https://doi.org/10.14236/ewic/HCI2018.40

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

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

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

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

Hearst, M. A. (2011). "Natural" search user interfaces. *Communications of the ACM*, *54*(11), 60–67. https://doi.org/10.1145/2018396.2018414

Horstmann, A. C., & Krämer, N. C. (2019). Great expectations? Relation of previous experiences with social robots in real life or in the media and expectancies based on qualitative and quantitative assessment. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00939

Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). *Communication and persuasion. Psychological studies of opinion change*. Yale University Press. http://psycnet.apa.org/record/1954-05952-000

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

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

Kinsella, B., & Mutchler, A. (2020). *Smart speaker consumer adoption report*. voicebot.ai. https://research.voicebot.ai/report-list/smart-speaker-consumer-adoption-report-2020/

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). Guilford Press.

Koh, Y. J., & Sundar, S. S. (2010). Effects of specialization in computers, web sites, and web agents on e-commerce trust. *International Journal of Human-Computer Studies*, *68*(12), 899–912. https://doi.org/10.1016/j.ijhcs.2010.08.002

Lankton, N., McKnight, D. H., & Thatcher, J. B. (2014). Incorporating trust-in-technology into Expectation Disconfirmation Theory. *The Journal of Strategic Information Systems*, *23*(2), 128–145. https://doi.org/10.1016/j.jsis.2013.09.001

Lankton, N., McKnight, D. H., & Tripp, J. (2015). Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, *16*(10), 880–918. https://doi.org/10.17705/1jais.00411

Lee, C.-J., Scheufele, D. A., & Lewenstein, B. V. (2005). Public attitudes toward emerging technologies: Examining the interactive effects of cognitions and affect on public attitudes toward nanotechnology. *Science Communication*, *27*(2), 240–267. https://doi.org/10.1177/1075547005281474

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, *46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50.30392

Lei, X., Tu, G.-H., Liu, A. X., Li, C.-Y., & Xie, T. (2018). The insecurity of home digital voice assistants—Vulnerabilities, attacks and countermeasures. *IEEE Conference on Communications and Network Security (CNS)*, 1–9. https://doi.org/10.1109/CNS.2018.8433167

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods*, *13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

Li, X., Hess, T. J., & Valacich, J. S. (2008). Why do we trust new technology? A study of initial trust formation with organizational information systems. *The Journal of Strategic Information Systems*, *17*(1), 39–71. https://doi.org/10.1016/j.jsis.2008.01.001

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The Media Are Social Actors paradigm. *Human-Machine Communication*, *2*, 29–55. https://doi.org/10.30658/hmc.2.2

Lyons, K. (2020, April 16). *Amazon's Alexa gets a new longform speaking style*. The Verge. https://web.archive.org/web/20200416230934/https://www.theverge.com/2020/4/16/21224141/amazon-alexa-long-form-speaking-polly-ai-voice

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709. https://doi.org/10.2307/258792

McKnight, D. H., Carter, M., Thatcher, J. B., & Clay, P. F. (2011). Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on Management Information Systems*, *2*(2), 1–25. https://doi.org/10.1145/1985347.1985353

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

McTear, M., Callejas, Z., & Barres, D. G. (2016). *The conversational interface: Talking to smart devices*. Springer International Publishing. https://doi.org/10.1007/978-3-319-32967-3

Metzger, M. J., Flanagin, A. J., Eyal, K., Lemus, D. R., & Mccann, R. M. (2003). Credibility for the 21st century: Integrating perspectives on source, message, and media credibility in the contemporary media environment. *Annals of the International Communication Association*, *27*(1), 293–335. https://doi.org/10.1080/23808985.2003.11679029

Moon, Y., Kim, K. J., & Shin, D.-H. (2016). Voices of the internet of things: An exploration of multiple voice effects in smart homes. In N. Streitz & P. Markopoulos (Eds.), *Distributed, ambient and pervasive interactions* (pp. 270–278). Springer International Publishing. https://doi.org/10.1007/978-3-319-39862-4_25

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

Natale, S. (2021). To believe in Siri: A critical analysis of voice assistants. In *Deceitful media: Artificial intelligence and social life after the Turing test*. Oxford University Press. https://doi.org/10.1093/oso/9780190080365.003.0007

Natale, S., & Cooke, H. (2020). Browsing with Alexa: Interrogating the impact of voice assistants as web interfaces. *Media, Culture & Society*, 0163443720983295. https://doi.org/10.1177/0163443720983295

Newhagen, J., & Nass, C. (1989). Differential criteria for evaluating credibility of newspapers and TV news. *Journalism Quarterly*, *66*(2), 277–284. https://doi.org/10.1177/107769908906600202

Newman, N. (2018). *The future of voice and the implications for news*. Reuters Institute for the Study of Journalism. https://web.archive.org/web/20211228171333/https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-11/Newman%20-%20Future%20of%20Voice%20FINAL_1.pdf

Newman, N., Fletcher, R., Kalogeropoulos, A., Levy, D. A. L., & Nielsen, R. K. (2018). *Digital news report 2018*. Reuters Institute for the Study of Journalism. https://web.archive.org/web/20190520173020/http://media.digitalnewsreport.org/wp-content/uploads/2018/06/digital-news-report-2018.pdf

Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. *Journal of Advertising*, *19*(3), 39–52. https://doi.org/10.1080/00913367.1990.10673191

Olson, C., & Kemery, K. (2019). *Microsoft voice report*. https://web.archive.org/web/20190502032155/https://about.ads.microsoft.com/en-us/insights/2019-voice-report

Oxford University Press. (n.d.). Hybrid. In *OED Online*. https://web.archive.org/web/20211228171746/https://www.oed.com/view/Entry/89809

Pertegal, M.-Á., Oliva, A., & Rodríguez-Meirinhos, A. (2019). Development and validation of the Scale of Motives for Using Social Networking Sites (SMU-SNS) for adolescents and youths. *PLOS ONE*, *14*(12), e0225781. https://doi.org/10.1371/journal.pone.0225781

Porter, J. (2019, January 16). *Alexa's news-reading voice just got a lot more professional*. The Verge. https://web.archive.org/web/20190116184000/https://www.theverge.com/2019/1/16/18185258/alexa-news-voice-newscaster-news-anchor-us-launch

Priest, S. H. (2001). Misplaced faith: Communication variables as predictors of encouragement for biotechnology development. *Science Communication*, *23*(2), 97–110. https://doi.org/10.1177/1075547001023002002

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

Rousseau, D. M., Sitkin, S. B., Burt, R. S., & Camerer, C. (1998). Introduction to special topic forum: Not so different after all: A cross-discipline view of trust. *The Academy of Management Review*, *23*(3), 393–404.

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Statista. (2021). *Global smart speaker vendors' market share 2016–2020*. https://web.archive.org/web/20210124041320/https://de.statista.com/statistik/daten/studie/818999/umfrage/marktanteile-der-hersteller-am-absatz-von-intelligenten-lautsprechern-weltweit/

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

Taş, S., Hildebrandt, C., & Arnold, R. (2019). *Sprachassistenten in Deutschland [Voice assistants in Germany]* (Discussion paper No. 441). Wissenschaftliches Institut für Infrastruktur und Kommunikationsdienste. https://web.archive.org/web/20211228172352/https://www.wik.org/uploads/media/WIK_Diskussionsbeitrag_Nr_441.pdf

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

Wang, W., & Benbasat, I. (2016). Empirical assessment of alternative designs for enhancing different types of trusting beliefs in online recommendation agents. *Journal of Management Information Systems*, *33*(3), 744–775. https://doi.org/10.1080/07421222.2016.1243949

Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung [Structural equation models]*. Springer-Verlag. https://doi.org/10.1007/978-3-642-35012-2

Yang, X., Aurisicchio, M., & Baxter, W. (2019). Understanding affective experiences with conversational agents. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1–12. https://doi.org/10.1145/3290605.3300772

Considering the Context to Build Theory in HCI, HRI, and HMC: Explicating Differences in Processes of Communication and Socialization With Social Technologies

Andrew Gambino[1] and Bingjie Liu[2]

1 Department of Communication Studies, Hong Kong Baptist University, Hong Kong
2 Department of Communication Studies, California State University, Los Angeles, CA, USA

Abstract

The proliferation and integration of social technologies has occurred quickly, and the specific technologies with which we engage are ever-changing. The dynamic nature of the development and use of social technologies is often acknowledged by researchers as a limitation. In this manuscript, however, we present a discussion on the implications of our modern technological context by focusing on processes of socialization and communication that are fundamentally different from their interpersonal corollary. These are presented and discussed with the goal of providing theoretical building blocks toward a more robust understanding of phenomena of human-computer interaction, human-robot interaction, human-machine communication, and interpersonal communication.

Keywords: human-computer interaction, human-robot interaction, interpersonal communication, social affordances, computers are social actors, message production

Introduction

Advances in computer technologies have resulted in the development of diverse and increasingly social technologies. Concurrently, we find these social technologies being adopted and used more frequently. As a product of these developments, social technologies facilitate

CONTACT Andrew Gambino ⓘ · Department of Communication Studies · School of Communication and Film · 936D Communication and Visual Arts Building · Hong Kong Baptist University · Kowloon Tong, Hong Kong · andrew.gambino@gmail.com

a growing amount of our communication, both as mediators (as in the case of computer-mediated communication, CMC) and as interactants (as in the case of human-computer interaction, HCI, and human-machine communication, HMC).

As communication scholars, the pace of technological advances, development of features and products, and their adoption present opportunities and challenges (Fortunati & Edwards, 2021; J. Fox & Gambino, 2021). On one hand, we are well-positioned to examine the social features that gain popularity with established methods and experience studying processes of communication both on- and offline. On the other hand, our research can be outpaced by developments in the modern technological landscape. To address this issue, we often focus our inquiries conceptually rather than technically through an affordance-based approach (Evans et al., 2016; Flanagin, 2020; J. Fox & Gambino, 2021; J. Fox & McEwan, 2017; Gambino et al., 2020; Rodríguez-Hidalgo, 2020; Sundar et al., 2015). An affordance-based approach engages with the concept underlying a feature or use, such as recordability or publicness, to establish a generalizable effect across features, platforms, and media. We strongly endorse the social affordance-based approach to the study of social technologies, and it is our aim in this manuscript to demonstrate the complementary utility of contextual factors of digital HMC through the lens of extant communication and social psychology theories.

Spence (2019) and Fortunati and Edwards (2021) have centrally positioned the question of "how" in the development of HMC theory. In limited space, Spence (2019) detailed avenues for the advancement of HMC theory. One such avenue, the process of applying existing theories from communication and related disciplines, is discussed, but framed as a separate endeavor from building "theories central to HMC" which Spence notes, "do exist and are being developed, tested and refined" (p. 286). We agree with both perspectives forwarded by Spence, particularly with respect to theory development (see: Edwards et al., 2019; Spence et al., 2014; Westerman et al., 2020). Here, we advance a third perspective that incorporates both the former and the latter. Namely, that communication scholars are uniquely positioned to build HMC theories through consideration of the relationship between contextual factors in HMC and those in theories of communication and relationships. With this perspective, communication scholars can both build theories that provide a more precise and comprehensive understanding of socio-technological phenomena and find answers to questions of extant theories of communication and relationships.

In this manuscript, we demonstrate the utility of our perspective through a discussion of two theoretical perspectives. Specifically, we explicate the processes of socialization and message production in the HMC context. We focus on the differences between the HMC and interpersonal context, advancing from description to theoretical potential. We first present our broader argument by engaging with processes of socialization, drawing heavily on social learning theory (Bandura, 1977, 1989). We then examine processes of interpersonal communication, focusing on contextually driven differences in goal structure that underlie message production (Dillard & Solomon, 2000).

Explicating the Context of HCI and HMC

Contextual factors play a central role in shaping how a communication episode unfolds (Dillard & Solomon, 2000). Communication is central in most human endeavors; as a

collective, the study of communication is multidisciplinary. We often define our research categorically, based on broad contextual differences in the persons (e.g., interpersonal communication, small-group communication) or the content of the communication act (e.g., health communication, political communication). HCI and HMC carry a bit of both, with differences in content (often digital or mediated) as well as differences in the interactant (Guzman & Lewis, 2020). Despite sharing qualities with other subfields, HMC is often considered a unique context. In practice, context is rarely explicated and is frequently used as a catch-all explanation for findings. As empirical studies have found both similarities (Edwards et al., 2020; Ho et al., 2018; S. K. Lee et al., 2021; Meng & Dai, 2021; Xu, 2019) and differences between HMC and interpersonal communication (Edwards & Edwards, 2022; Jia et al., 2022; Kim & Song, 2021; Liu & Wei, 2021; van Straten et al., 2021; van Straten at al., 2022; Zellou et al., 2021), we argue that explication of the HMC context in relation to the interpersonal corollary is necessary to provide meaningful and nuanced explanations for groups of findings, and, ultimately, to build theories of both HMC and interpersonal communication.

Following Dillard and Solomon (2000), we demonstrate how major contextual factors of *digital* HMC (i.e., communication with a computer or computer-powered agent, inclusive of physical, virtual, and any hybrid-type forms) can be utilized to engage with extant theories. As Guzman and Lewis (2020) suggested, central to HMC is the nonhuman nature of the interactant or partner. Fox and Gambino (2021) engaged with the implications of humanoid social robots and how their varying levels of social affordances relate to relational development theories. For example, they suggested the inappropriateness of applying social penetration theory (Altman & Taylor, 1973) to relationships between humans and social robots, as social robots have limited abilities to retain (i.e., persistence) and recall (i.e., searchability) prior interactions. These deficiencies make it difficult to have in-depth conversations with social robots, and deep conversations are considered essential to the development of relationships (Altman & Taylor, 1973).

In this manuscript, we present context as an integral consideration to theory development. We demonstrate the usefulness of such considerations through the examples of socialization and message production. In the case of long-term socialization, we focus our discussion on social learning theory and how consideration of the digital context provides ground for novel inquiries. In the case of message production, considering the context of digital HMC both provide a novel method of studying the interpersonal phenomenon of goal structures and allow for a valid assessment of the specific aspects of the framework relevant to the conditions of HMC.

Social Learning Theory: Observational and Experiential Learning in HMC

Researchers in HCI and digital HMC understand a critical contextual factor: computers are not human, and most people believe computers do not require or deserve social treatment (Reeves & Nass, 1996). We see the dichotomy of social and nonsocial, however, to be a bit limited for a comprehensive understanding of the modern technological context. In studies, researchers can cite differences in perceptions of sociability or some specific but nongeneralizable differences between a human and a computer as explanations for their findings. For

example, heuristics (i.e., mental shortcuts) that are formed within systems of social cognition (Fiske & Taylor, 1984) are frequently invoked to explain differences between humans and media agents (e.g., Gambino et al., 2019). Often, differences between perceptions of humans and digital interlocutors are attributed to such heuristics that, in practice, account for groups or sets of features and affordances that likely differ between the objects of study. In this manuscript, we aim to provide researchers with an approach that allows for the construction of theories unique to digital HMC, as well as a path to improve the precision of extant theories through the study of social technologies. We first discuss the larger process of socialization through the lens of Bandura's social learning theory (1971).

Social learning theory was developed by Albert Bandura (1971; Bandura & Walters, 1963), building on the work of Robert Sears (1957). Although both suggest that persons develop through social interactions, Bandura and Walters (1963) noted the importance of observational learning and the role it could play in the acquisition of a person's behavioral practices and norms (Grusec, 1992). The observational learning component is perhaps best exemplified by the "bobo doll" studies (Bandura, 1965; Bandura et al., 1961). In Bandura et al. (1961), children demonstrated more aggressive behaviors after watching aggressive behaviors. Additionally, Bandura (1965) demonstrated that the tendency to engage in aggressive behaviors is influenced by whether the model's aggressive behaviors were rewarded or punished, a phenomenon known as *vicarious reinforcement*. As predicted, children that saw aggressive behaviors rewarded were more likely to behave aggressively, whereas children that saw aggressive behaviors punished were less likely to behave aggressively. Through observational learning and vicarious reinforcement, Bandura demonstrated that people could learn both an act and its consequences through observation.

Jesse Fox and her colleagues have demonstrated that observational learning and vicarious reinforcement processes occur also within virtual environments (J. Fox & Bailenson, 2009; J. Fox et al., 2013; J. A. Fox, 2010). For example, J. Fox and Bailenson (2009) found participants exercised more after watching their avatar experience reward (shrinking in size) and punishment (growing in size). Notably, in their second study, J. Fox and Bailenson (2009) found that these effects were more pronounced when the representations resembled the participant. As J. Fox and Bailenson (2009) suggested, virtual environments provide new ground to study identification; the form a person can take in a virtual environment, either volitionally or non-volitionally, is qualitatively different, and greater in number, than in our natural setting.

In addition to social learning occurring within virtual environments, here we consider the implications for social development that occurs within digital HMC but outside of the virtual environment. Consider the use of a computer. Although we can infer social elements within a virtual environment to construct our models for navigating social environments (Walther, 1992; Walther, 1996), direct observation or engagement of acts and outcomes are restricted. In physical environments, people can more easily observe social episodes. For example, children can observe bullying on a playground without being directly involved as the bully or the bullied person. Within a virtual environment, such an act is observable by the user of the computer, but for a third party, such as a child observing his or her parent using the computer, these acts and their consequences are less observable because of physical and digital barriers or boundaries. Related to social learning, rewards and punishments may be understood by the engaged user, but they are less likely to be understood

by third parties through traditional processes of observational learning. These device-induced boundaries are examined by researchers in developmental psychology. For example, researchers have found that parental engagement with social technologies can bring about relational barriers with their children (For review: Hassinger-Das et al., 2020). In an experiment, Gaudreau et al. (2021) found that children asked fewer questions to phone-engaged parents while the parent-child dyad attempted to complete a novel task. Furthermore, the phone-engaged parents asked their children fewer information-seeking questions, which suggests that engagement with personal computers, such as mobile phones, may influence adult behavior as well. Altogether, we argue that there is less to be learned from observing a peer or parent communicate through or with a computer.

As discussed, within virtual environments these boundaries are mitigated by the presence of other persons, but in communication with machines they remain, even if tempered by the collective use of a given machine. For example, although interactions with virtual assistants (e.g., Alexa) occasionally occur in the presence of other persons, many will occur in private, and these interactions may be lacking the social stakes necessary for vicarious reinforcement to occur.

A lower frequency of observational learning and vicarious reinforcement in digital HMC has implications for social development. If nothing else, it may lead to a developmental calculus (i.e., the ratio of experiential and observational learning) that relies more heavily on experiential learning. With a considerable amount of digital HMC occurring experientially and with different consequences, it occurs to us that the scripts, models, or schema shaped through these digital HMC interactions and relationships may be quite personal.

Significant empirical work is necessary to assess the validity of the preceding claims, as well as any connections made to data. With that said, a developmental calculus that relies less on traditional processes of socialization may result in different behavioral norms. Consider, as an example, aggression, particularly a person's verbalized acts when interacting with virtual agents. Researchers have observed a high rate of verbal aggression in users' commentary toward machine agents (10% of language considered abusive in De Angeli & Brahnam, 2008). Among 59 adolescents that interacted with a female conversational pedagogical agent, nearly 40% of the students were aggressive toward the agent with hyper-sexualized and dehumanized commentary such as "shut up u hore" and "want to give me a blow job" (Veletsianos et al., 2008). In the context of online chatting, research has found individuals were less open, agreeable, extroverted, and conscientious when interacting with a chatbot as reflected in their messages than when interacting with a human friend initially (Mou & Xu, 2017).

On their face, these findings are disturbing, and when interpreted at a macro level they may reflect threats to our interpersonal relationships. We suggest three context-based explanations for these behaviors, all of which draw on social learning theory. For one, the increased use of socially inappropriate language may be an outcome of observational learning. Although we have argued that there are fewer opportunities for observational learning to occur, it may make such acts more important to the third party. In that sense, due to a lack of diverse interactions, a parent or peer that demonstrates aggression, frustration, or communicates to nonhuman agents with dehumanizing language may have an out-sized influence on the observer. Second, such communication patterns may be the result

of increased experiential learning and more personalized scripts. Third, the lack of human presence in digital HMC, regardless of perceptions of social presence, may bring about interactions that provide fewer considerations of interpersonal norms. In private use, there may be less anchoring to social norms and practices that might occur during collective use of a technology (e.g., when families or friends use Alexa in the same room). Although subversion can be a positive, there are desirable interpersonal acts that may be more difficult to find, or even present, in a more saturated HMC environment. If such behaviors are the product of an updated calculus of experiential and observational learning in digital HMC, the scripts or schema being formed through digital HMC may be problematic, especially if there is an interplay between one's interpersonal scripts or schema for communication or relationships, as researchers have observed in virtual environments (Velez et al., 2019; Velez et al., 2021).

We turn now to the process of message production, with our focus squarely on how theories of interpersonal communication can be engaged through consideration of the digital HMC context.

Message Production and Goal Structure as Outcomes of the Digital HMC Context

Communication is context dependent. When our communication partners are digital machines, rather than humans, how do we change our processes of message production? To answer this question, it is beneficial, if not imperative, to understand the impact of context on message production. Originating from the view of message production as a goal-driven process (for a review, see Meyer, 2021), and in an effort to provide a more comprehensive and useful understanding of the message production process, Dillard and Solomon (2000) conceptualized communication context "in terms of perceived empirical regularities in social reality (i.e., social densities) and the configurations of interpersonal goals that follow from them (i.e., goal structures)" (p. 167).

Interpersonal communication researchers generally differentiate between two goals in message production: primary goals, which are the primary reason for producing a message and define the meaning of the interaction, and secondary goals, which are concerns that arise from considering how to achieve the primary goal and thereby constrain the interaction (Dillard et al., 1989). For example, while pursuing a primary goal, one may also need to consider secondary goals, such as maintaining their identities or relationships with their partner (Clark & Delia, 1979; Dillard et al., 1989). A person may also pursue secondary goals related to linguistics, such as clarity and relevance in their speech (Hample & Dallinger, 1987), especially when the environment and the pursuit of the primary goals heighten such concerns. The configurations of primary and secondary goals constitute goal structures, which arise in given social densities (Schrader & Dillard, 1998).

Social densities are configurations of the obstacles and opportunities to engage in certain behaviors that people perceive in their social reality (Dillard & Solomon, 2000). Goals arise within social densities such that people tend to form goals that are possible or allowed for by the social reality (Heckhausen & Kuhl, 1985), and may also try to overcome the obstacles in cases where the obstacles hinder their goal achievement (Roloff & Janiszewski, 1989). In other words, the anticipated obstacles and opportunities a message producer perceives

in a given situation shapes their goal structure. Hence, if the context of digital HMC is constituted by social regularities that are qualitatively or quantitatively different from their interpersonal corollary, it follows that different goal structures should arise and, ultimately, different messages will be produced.

Empirical research has corroborated the proposition that goals both arise within our social reality and are facilitated and constrained by it; certain goals are more likely to be triggered in the presence of certain situational factors (e.g., Dillard et al., 1989; Hample, 2016). For example, in an experiment (Hample, 2016, Study 2), participants were randomly assigned to scenarios with different settings (topics, places, relationship types, sex of the partner) and were instructed to imagine a conversation that would follow. By examining the goals that participants reported in the imagined conversations, the researchers found that an academic setting is less likely to trigger the primary goal of relationship maintenance. When at least one of the interactants was a woman, though, relationship maintenance was more likely to be the primary goal pursued. As explained from the perspective illustrated in Dillard and Solomon (2000), academic settings are more likely to be associated with factors that facilitate professional, rather than personal, activities (i.e., social densities). Hence, relational goals may be discouraged. When the interactant is a woman, however, the stereotype that women have more interest and expertise in relational issues (i.e., social densities) may facilitate the pursuit of relational goals.

In another study (Schrader & Dillard, 1998), participants reported the perceived importance of primary and secondary goals within 1 of 15 social episodes; it was found that the primary goal of persuasion (i.e., to change attitudes and behaviors) often revolves around close relationships and is associated with heightened concerns for social appropriateness or politeness (i.e., a secondary goal). Similarly, Wilson et al. (1998) found that the primary goal of giving advice often prompts a secondary goal of managing or maintaining the relationship with the target. This is because human partners have the desire to maintain face; that is, being appreciated while not being impeded by others. The attempt to change others' opinions or behaviors may threaten others' face (Brown & Levinson, 1978). Within such social regularities, the face-threatening goals should be less inhibited within close relationships where face protection is of lesser concern; whereas when a primary goal of persuasion must be pursued outside close relationships, it should be pursued with concerns of face protection and relationship maintenance to avoid reactance and undesired relational outcomes.

In sum, contextual factors influence message production and design by dictating the primary and secondary goals. As such, variations of social affordances and broader contextual factors of digital HMC may trigger or inhibit certain goals, leading to differences in both the content of the primary and secondary goals, as well as the overall complexity of the goal structure. Next, we discuss how such goals and goal structures intersect with contextual factors of digital HMC.

Goal Structures in HMC and Implications for HMC Research

Primary Goals in HMC

According to Hample's theory of interpersonal goals and situations (2016), primary goals are influenced by obvious factors in the situation ranging from settings and topics to salient

characteristics of the target. Given the unique features of machines, individuals may pursue different primary goals in digital HMC, as compared with interpersonal, human communication.

Proper conceptualization of goals in digital HMC requires consideration of core tenets of goals in communication. For example, in most interpersonal communication situations, humans are producers or receivers (Shannon & Weaver, 1949). This remains true in our modern technological landscape, and computers most often serve as a mediator (CMC). Advances in technologies have complicated this formula. Computers are often now considered distinct sources (Sundar & Nass, 2000), and, in HMC, researchers have suggested that computers can serve as active receivers as well (Guzman & Lewis, 2020).

Computers have traditionally been designed to help users achieve specific goals. Because of this, the communication settings or topics in digital HMC are often defined by the functions of the computers or machines, which then stipulates primary goals accordingly. In other words, an individual's primary goals in digital HMC are not always emergent, but they are often defined and confined within the range of that person's perception of the computer's capacities. Because of certain capacities that machines have or lack compared to humans, either in actuality or perceived by the user, certain primary goals might be more or less frequently pursued in digital HMC than in human communication. We frame our discussion in light of commonly held perceptions and beliefs about the capacities of computers and media agents (e.g., Gray et al., 2007; Sundar & Kim, 2019).

According to Sundar (2008), machines are believed to have less fallible "memory" and are capable of gathering and processing larger amounts of data than humans. Consistent with this suggestion, many machines are designed to serve as cognitive aids for humans. For example, many digital technologies provide factual information, keep records, or set reminders, amongst other similar acts; and users frequently interact with such machines driven primarily by goals of cognitive assistance, such as seeking accurate information (Hamilton et al., 2016).

Another unique aspect of machines, compared with human counterparts, is their heightened degree of agreeability. Machines such as robots and chatbots are designed to obey and accept users unconditionally. An analysis of user reviews of Replika, a chatbot that is designed to be a person's companion, showed that users mentioned the constant positivity that Replika demonstrated, with comments such as "It always gives me compliments and cheers me up," making them feel loved and accepted (Ta et al., 2020). Although unconditional positivity is not always a healthy thing, this characteristic of machines, by providing the opportunities for more positive conversations, may then facilitate people to initiate interactions with machines to obtain social support, especially esteem support.

Compared with humans, digital machines have a poorer understanding of complex human languages and the social context of human affairs. As J. Fox and Gambino (2021) argued, few, if any, digital machines are designed, or have the capacity, to understand complex social contexts and to have sophisticated, personal conversations with human users, which are capacities essential for the development of genuine, two-way relationships. For example, in interviews with users of Replika, they reported frustrations in its lack of ability to understand complex social contexts and norms such as how frequently to bring up a user's ex (Skjuve et al., 2021). As a result, people should be less likely to pursue goals that involve soliciting deep understanding or developing meaningful relationships with

digital machines. As Mou and Xu (2017) found in their study comparing messages produced during initial conversations with chatbots or people, participants engaged in less self-disclosure with chatbots, which is considered the key to building relationships (Altman & Taylor, 1973).

Secondary Goals in HMC

Because machines lack experiential capacities, such as emotions (Gray et al., 2007), as well as the ability to make social judgments (Sundar & Kim, 2019), secondary goals pertaining to such capacities that are common in interpersonal communication should be less relevant in HMC. Examples of such secondary goals include, but are not limited to, avoiding face threats, relationship maintenance, and impression management (Meyer, 2009). These secondary goals are premised on the target's inner experience (e.g., face, social judgments, well-being), which machines lack (Gray et al., 2007). Therefore, they should be activated less frequently during digital HMC.

Although this argument has not been systematically examined, empirical findings suggest its plausibility. First, the lack of concern for the "feelings" of a machine is supported by the high rate of verbal aggression observed in HMC, such as verbal aggression toward machines (De Angeli & Brahnam, 2008; Veletsianos et al., 2008). In addition, researchers have also found that aggression is especially high when less mind is attributed to a machine agent (Keijsers & Bartneck, 2018), which suggests that lower perceptions of mind in machines might suppress secondary goals of protecting a target's feelings or maintaining the relationship while users pursue primary goals such as obtaining information from machines.

Second, people have reported less concerns of social consequences, or social judgments, during digital HMC. The lack of social concerns in digital HMC is reflected in the high sensitivity of persons' self-disclosure to digital machines (Brandtzaeg & Følstad, 2018; Kretzschmar et al., 2019; Ta et al., 2020). For example, in an experiment where participants were interviewed by either a faceless computer system or a human, participants disclosed more sensitive information, with greater detail, to the computer interviewer (Pickard & Roster, 2020). Similarly, Skjuve et al. (2021) found that many users of Replika moved quickly to disclosure on personal, intimate topics; skipping the phase of "orientation" (e.g., having superficial small talks), which is considered the first stage of interpersonal communication for people to get to know each other and establish initial trust according to social penetration theory (Altman & Taylor, 1973). In their study, some users elaborated on this phenomenon of moving quickly to personal or intimate disclosures, and the researchers described it as "they [participants] did not see any social risks in this sharing given Replika's non-judgmental character" (Skjuve et al., 2021, p. 5). Skjuve et al. also found that participants self-disclosed to Replika information that they would not typically feel comfortable disclosing to a human, due to concerns of social norms, such as personal problems and sexual orientation.

In another study where participants interacted with a chatbot designed to make small talk, researchers found the chatbot induced deep self-disclosure from participants during 3 weeks of use (Y.-C. Lee et al., 2020). In their follow-up interviews, participants expressed how carefree they were when answering the chatbot's sensitive questions, often making reference to the nonjudgmental or feelingless nature of the chatbot. For example, one

participant said, "I can say anything to the chatbot. If I'm texting with an anonymous online person, I still cannot disclose everything. I would think about the person's feelings and how s/he would react" (p. 7). Another said,

> The chatbot once asked me about a sexual relationship . . . Because [the] chatbot is not a human, I don't feel embarrassed. I know that there is a research team behind the chatbot, but I'm facing only the chatbot when giving my answers, and feel safe doing so. (p. 7)

Collectively, these findings suggest that the goal of impression management, which is often associated with the primary goal of self-disclosure in human communication (Meyer, 2009), may be suppressed or invoked less frequently in digital HMC, which then facilitates more carefree self-disclosure with machines (Brandtzæg et al., 2021).

Compared to humans, machines are less subject to biological constraints. For example, a machine does not need rest and a machine can be mass-produced and distributed to multiple users simultaneously without compromising its performance, though machines often require maintenance and human resources to maintain a level of performance. Due to biological factors, digital machines are less constrained by time and space than humans and they can be more mobile and can produce, or work, without physical or mental constraints. Therefore, individuals engaging in digital HMC should have less concern about social appropriateness related to digital machines' availability in terms of time and space and less concern about burdening or inconveniencing a machine partner, because a machine does not endure any loss by allocating time and "attention" to a user. By contrast, when interacting with a human partner, the appropriateness of time and location to initiate a conversation does matter for the communication processes and consequences.

These propositions are supported by empirical findings. For example, users of Replika reported that when they needed to talk (e.g., when they were stressed), they would go to Replika because it would not matter if they were on the bus or at a restaurant, and they did not need to bug their potentially busy friends. Further, users explicitly mentioned that Replika had qualities that humans do not have, such as being available at all times and therefore easier to open up to (Skjuve et al., 2021). In the user reviews of Replika, its availability was frequently mentioned as one of its merits (Ta et al., 2020). In another study of Woebot, a chatbot for mental health, users specifically mentioned and valued its nature of being unconstrained by time and space, allowing them to potentially just sit on the subway or in their room and receive informational support (Bae Brandtzæg et al., 2021).

Such qualities of digital machines may suppress or deactivate goals related to selecting an appropriate time and space to initiate a conversation, and therefore reduce the cognitive burden of people. This may, in turn, lead people to pursue their primary goals with machines more frequently than with humans. With that said, social technologies vary in terms of their locatability and portability (Rodríguez-Hidalgo, 2020; Schrock, 2015). For example, social robots may be less flexible (e.g., harder to move, requiring electricity to operate) than chatbots, and therefore social robots may activate secondary goals related to time and space more frequently.

In addition to suppressing these secondary goals common in human communication, digital HMC may trigger novel secondary goals that are not frequently considered during common acts of human communication. First, because machines are deficient in their ability to understand and contextualize human communication, secondary goals related to understandability or efficiency may be triggered. For example, Muresan and Pohl (2019) found that users of Replika reported limitations in its conversational capabilities, and users were therefore concerned about the degree to which a machine would understand them.

Second, digital machines are often high in recordability (e.g., digital or digitized messages are often stored in a database). Such a high level of recordability may trigger concerns of privacy and confidentiality when the communication involves the disclosure of personal or sensitive information. This may lead persons to consider a secondary goal of information protection, leading to less breadth and depth in self-disclosure. For example, some Replika users have reported that they were concerned about how it would manage their private information; to address these concerns, they investigated Replika's terms on privacy and information security, asked Replika about how their data would be stored and who would have access to it, or contacted the provider to request such information when they wanted to disclose private information to Replika (Skjuve et al., 2021). Additionally, when designed with more transparent privacy policies for data processing and storage, robots provide a better user experience (Vitale et al., 2018).

Goal Structure and the Interplay Between Primary and Secondary Goals in HMC

In human communication, "primary goals bring about secondary goals" (Dillard & Solomon, 2000, p. 171). That is, primary goals are found to be reliably associated with certain secondary goals as discussed (e.g., Schrader & Dillard, 1998; Wilson et al., 1998). In digital HMC, however, these structures, or common configurations, may be affected by contextual factors of the interactant and relationship. As we have argued, concerns for the target's well-being, the speaker-target relationship, and social judgments on the speaker, which are common secondary goals in human communication, are likely to manifest in qualitatively or quantitatively different forms in digital HMC. For example, to solicit help from a digital machine or to disclose a personal failure need not always trigger concerns for one's own face, a secondary goal of impression management that is frequently triggered in the interpersonal corollaries (Meyer, 2009; Wilson et al., 1998). As a result, differences, in terms of both the complexity and the content of goal structure, may exist between digital HMC and human communication. Specifically, we expect a simpler goal structure in HMC (i.e., fewer secondary goals) as compared with interpersonal, human communication.

As for the content of the goal structure, we may also observe different configurations of goal structures in digital HMC. Compared to the human communication corollary, certain primary goals may pair more or less frequently with secondary goals. For example, primary goals that involve complex problem-solving (e.g., obtaining support for a malfunctioning product) may trigger concerns of a machine's ability to understand and contextualize language or to respond in a contingent manner. These reflect secondary goals that are likely to be less relevant or less frequent in conversations with a human agent.

Implications for Processes of Communication and Socialization

Message production is goal-driven. Understanding the context of digital HMC in terms of individuals' goal structures and broader processes of socialization allows us to make insightful and focused inquiries into the processes and outcomes of digital HMC. In the remainder of this manuscript, we discuss the micro and macro level implications of considering processes of socialization and individuals' goal structures for research in HMC. We close the manuscript with a discussion on research practices, encouraging our colleagues to consider the importance of context to increase the internal validity of our work.

The differences in individuals' goal structures when they interact with computers and humans may explain the differences between HMC and human communication observed, such as less intimacy and self-disclosure (Mou & Xu, 2017) and high aggression (De Angeli & Brahnam, 2008; Strait et al., 2017; Veletsianos et al., 2008) in HMC. Alternatively, or concurrently, a deficit in observational learning may lead to such outcomes as a person develops personalized scripts for use that encourage aggression, without punishment, and seeks and finds fewer rewards for intimacy or self-disclosure.

Further, we argue that differences in the content, complexity, and configurations of goal structures in digital HMC and human communication may have downstream effects on human communication and, in the long run, our social skills. For example, soliciting cognitive aid is a prominent goal in digital HMC, and we can reasonably expect to observe a large number of digital HMC interactions to be question-answer type prompts. With the primary and secondary goals of relational maintenance and development suppressed in such interactions, we expect to observe fewer relational talks, discussions on complex issues, and lower self-disclosure depth which may result in a person developing less social-emotional skills, particularly those related to narratives and emotions. Because we have limited time and abilities to communicate, when we consider the socialization process, this allows us to see a more dire outcome of a person's engagement with digital interlocutors. Although such effects may not be observable in a single, or even across multiple, studies, as we have argued, there is theoretical justification for the consideration of how repeated engagement in less personal, sophisticated communication may influence one's interpersonal expectations and behaviors.

On a more micro level, with considerations for a machine as a social other, such as the machine's well-being, judgments, face, availability (i.e., secondary goals) suppressed, we expect higher directness and lower politeness in messages sent to machines than to humans under the same primary goals. We also expect people to engage in digital HMC with fewer temporal and spatial constraints.

When the secondary goal of being understood is triggered, we expect messages in HMC reflect more effort of accommodation; the use of less complex words and sentences, more context-independent messages, and more paraphrasing when compared to interpersonal communication. Additionally, if the secondary goal of privacy management is triggered, messages of self-disclosure in HMC may contain less private information. For example, individuals may be less willing to disclose information that may incriminate them to a therapist chatbot than to a human therapist, as a computer therapist is likely to have a digital record of the disclosure.

As we have argued, individuals may have a simplified goal structure in digital HMC, due to the suppression of many secondary goals common in human communication. A corollary to interpersonal communication suggests that individuals of low cognitive complexity or low cognitive resources may handle digital HMC better than human communication. Following, they may choose to achieve goals through digital HMC over human communication, or they may choose to engage more frequently with computers. This could explain why people prefer humans vs. machines for a certain task, as well as lead to the development of personalized scripts for interactions with computers.

Although our discussion so far focuses on the distinctive qualities of HMC context, similarities between interpersonal communication and digital HMC do exist, and in explicating context, it is worthwhile to note these as well. For example, Berger's plan-based theory of strategic communication (Berger, 1995, 1997) suggests that individuals will first use the available plan in long-term memory, but once frustration occurs, may seek alternative communication plans. In cases when the human-machine differences are irrelevant for the achievement of certain goals so that no frustrations occur, we expect similarities between interpersonal communication and digital HMC when individuals will apply well-learned, interpersonal scripts effectively.

Implications for Theory Development

There remains considerable value in findings that demonstrate similarities and differences between humans and machines (Edwards & Edwards, 2022; Ho et al., 2018). It is only through acknowledging differences and similarities at the larger level that more focused theoretical inquiries can be organized and examined. Now, we suggest moving forward with the study of computers and digital machines, focusing less on the global differences between humans and machines in general and more on the social affordances provided and enacted through HMC (Liu, 2018, 2021). Although in our previous discussion we made the contrast between HMC and human communication, we have discussed the goal structures in HMC in light of specific machine affordances and the lack of thereof. Social technologies vary in each particular form, and machines are developed in relation to the impossible to be known advances in underlying technologies of the future; thus, we recommend that predictions on the goal structures and communication with machines should be made with consideration of specific machine affordances.

Considering the differences between the goal structures in digital HMC and human communication, we argue there are several limitations in the current paradigms of HMC research and point out some alternative directions. In many digital HMC studies, participants are instructed to communicate with machines in a given context, with primary goals predefined or specified by the researchers, and then asked to evaluate the interaction experience and the machine agent (e.g., Edwards et al., 2020; Liu & Sundar, 2018). Considering the effects of machine affordances on primary goals, while also considering the effects of primary goals on secondary goals (Hample, 2016), such a research paradigm may artificially induce two antecedents for communication that might be nonexistent in natural settings: the goal that researchers stipulate for the interaction and an evaluation goal, which may distort the processes of digital HMC. These threats to ecological and external validity take on additional weight when considered with the arguments in this manuscript.

As empirical researchers, we are often trained to purposefully ignore or downplay threats of ecological and external validity to maximize internal validity, but without appropriate explication and consideration of the digital HMC context, we believe we may be adding artificiality to a naturally occurring process of digital HMC, therefore increasing threats to the internal validity of the findings. The similarities observed between human communication and digital HMC might be due to such an artifact of forcing primary goals to be identical, when, in actuality, the goals would have not been the same in the first place. To mitigate this threat, researchers may consider recording participants' goals in natural settings with methods such as experience sampling or surveys. Additionally, more clarity on the context of digital HMC may emerge through methods such as interviews or diary studies. We encourage researchers to undertake these methods with consideration of the digital HMC context situated both within and against existing theories, so that these methods may be used to provide a rich and focused description of HMC. We hope that with such understanding, deductive methods such as experiments can be employed to test causal relationships with less threats to both internal and external validity.

Conclusion

Through consideration of the digital HMC context, we have situated empirical findings and adjusted theoretical propositions of human communication and socialization. In these reconfigurations, we see benefits to communication scholars as our perspectives on social learning theory and goal structures focus directly on acts of communication. We look to the future of research in communication with hope, and we present these propositions for empirical testing, but also as examples of the means to theoretical engagement. We do not consider this manuscript to present anywhere close to a comprehensive integration of contextual factors of digital HMC or theories of communication and socialization. Instead, we encourage our colleagues to take these as examples and dive into theoretical spaces where their interests lie.

Acknowledgments

Author Biographies

Andrew Gambino (PhD, The Pennsylvania State University) is an Assistant Professor in the Department of Communication Studies at Hong Kong Baptist University. Andrew uses empirical methods to study communication, relationships, and technology. In his research, he examines how communication through and with social technologies both affects and reflects our interpersonal relationships.

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

Bingjie Liu (PhD, The Pennsylvania State University) is an Assistant Professor in the Department of Communication Studies at California State University, Los Angeles. She studies the nature of humanity manifested in human communication with technologies.

🆔 https://orcid.org/0000-0001-9019-8521

References

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

Brandtzæg, P. B., Skjuve, M., Kristoffer Dysthe, K. K., & Følstad, A. (2021). When the social becomes non-human: Young people's perception of social support in chatbots. In *Proceedings of the SIGCHI '21 Human Factors in Computing Systems* (CHI '21, pp. 1–13). https://doi.org/10.1145/3411764.3445318

Bandura, A. (1965). Influence of models' reinforcement contingencies on the acquisition of imitative responses. *Journal of Personality and Social Psychology*, *1*(6), 589–595. https://doi.org/10.1037/h0022070

Bandura, A. (1971). *Social learning theory*. General Learning Press.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

Bandura, A. (1989). Human agency in social cognitive theory. *American Psychologist*, *44*(9), 1175–1184. https://doi.org/10.1037/0003-066X.44.9.1175

Bandura, A., Ross, D., & Ross, S. A. (1961). Transmission of aggression through imitation of aggressive models. *The Journal of Abnormal and Social Psychology*, *63*(3), 575–582. https://doi.org/10.1037/h0045925

Bandura, A., & Walters, R. H. (1963). *Social learning and personality development*. Holt Rinehart, & Winston.

Berger, C. R. (1995). A plan-based approach to strategic communication. In D. E. Hewes (Ed.), *The cognitive bases of interpersonal communication* (pp. 141–179). Erlbaum.

Berger, C. R. (1997). *Planning strategic interaction: Attaining goals through communicative action*. Erlbaum.

Brandtzaeg, P. B., & Følstad, A. (2018). Chatbots: Changing user needs and motivations. *Interactions*, *25*(5), 38–43. https://doi.org/10.1145/3236669

Brown, P., & Levinson, S. (1978). Universals in language usage: Politeness phenomena. In E. Goody (Ed.), *Questions and politeness: Strategies in social interaction* (pp. 56–289). Cambridge University Press.

Clark, R. A., & Delia, J. G. (1979). Topoi and rhetorical competence. *Quarterly Journal of Speech*, *65*(2), 187–206. https://doi.org/10.1080/00335637909383470

De Angeli, A., & Brahnam, S. (2008). I hate you! Disinhibition with virtual partners. *Interacting with Computers*, *20*(3), 302–310. https://doi.org/10.1016/j.intcom.2008.02.004

Dillard, J. P., Segrin, C., & Harden, J. M. (1989). Primary and secondary goals in the production of interpersonal influence messages. *Communications Monographs*, *56*(1), 19–38. https://doi.org/10.1080/03637758909390247

Dillard, J. P., & Solomon, D. H. (2000). Conceptualizing context in message production research. *Communication Theory*, *10*(2), 167–175. https://doi.org/10.1111/j.1468-2885.2000.tb00186.x

Edwards, A., & Edwards, C. (2022). Does the correspondence bias apply to social robots?: Dispositional and situational attributions of human versus robot behavior. *Frontiers in Robotics and AI*, *8*, article 788242. https://doi.org/10.3389/frobt.2021.788242

Edwards, A., Edwards, C., & Gambino, A. (2020). The social pragmatics of communication with social robots: Effects of robot message design logic in a regulative context. *International Journal of Social Robotics*, *12*, 945–957. 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

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

Fiske, S. E., & Taylor, S. T. (1984). *Social cognition: From brains to culture*. Random House.

Flanagin, A. J. (2020). The conduct and consequence of research on digital communication. *Journal of Computer-Mediated Communication*, *25*(1), 23–31. https://doi.org/10.1093/jcmc/zmz019

Fortunati, L., & Edwards, A. (2021). Moving ahead with human-machine communication. *Human-Machine Communication*, *2*, 7–28. https://doi.org/10.30658/hmc.2.1

Fox, J., & Bailenson, J. N. (2009). Virtual self-modeling: The effects of vicarious reinforcement and identification on exercise behaviors. *Media Psychology*, *12*(1), 1–25. https://doi.org/10.1080/15213260802669474

Fox, J., Bailenson, J. N., & Tricase, L. (2013). The embodiment of sexualized virtual selves: The Proteus effect and experiences of self-objectification via avatars. *Computers in Human Behavior*, *29*(3), 930–938. https://doi.org/10.1016/j.chb.2012.12.027

Fox, J., & Gambino, A. (2021). Relationship development with humanoid social robots: Applying interpersonal theories to human-robot interaction. *Cyberpsychology, Behavior, and Social Networking*, *24*(5), 294–299. https://doi.org/10.1089/cyber.2020.0181

Fox, J., & McEwan, B. (2017). Distinguishing technologies for social interaction: The perceived social affordances of communication channels scale. *Communication Monographs*, *84*(3), 298–318. https://doi.org/10.1080/03637751.2017.1332418

Fox, J. A. (2010). *The use of virtual self models to promote self-efficacy and physical activity performance* [Unpublished doctoral dissertation]. Stanford University.

Gaudreau, C., Hirsh-Pasek, K., & Golinkoff, R. M. (2021). What's in a distraction? The effect of parental cell phone use on parents' and children's question-asking. *Developmental Psychology*, advance online publication. https://doi.org/10.1037/dev0001268

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

Gambino, A., Kim, J., & Sundar, S. S. (2019). Digital doctors and robot receptionists: User attributes that predict acceptance of automation in healthcare facilities. In *Proceedings of SIGCHI '19 Human Factors in Computing Systems* (CHI '19; LBW0287). https://doi.org/10.1145/3290607.3312916

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science*, *315*(5812), 619. https://doi.org/10.1126/science.1134475

Grusec, J. E. (1992). Social learning theory and developmental psychology: The legacies of Robert Sears and Albert Bandura. *Developmental Psychology*, *28*(5), 776–786. https://doi.org/10.1037/0012-1649.28.5.776

Guzman, A. L., & Lewis, S. C. (2020). Artificial intelligence and communication: A human–machine communication research agenda. *New Media & Society*, *22*(1), 70–86. https://doi.org/10.1177/1461444819858691

Hamilton, K. A., McIntyre, K. P., & Hertel, P. T. (2016). Judging knowledge in the digital age: The role of external-memory organization. *Applied Cognitive Psychology*, *30*(6), 1080–1087. https://doi.org/10.1002/acp.3277

Hample, D. (2016). A theory of interpersonal goals and situations. *Communication Research*, *43*(3), 344–371. https://doi.org/10.1177/0093650214565899

Hample, D., & Dallinger, J. M. (1987). Individual differences in cognitive editing standards. *Human Communication Research*, *14*(2), 123–144. https://doi.org/10.1111/j.1468-2958.1987.tb00124.x

Hassinger-Das, B., Brennan, S., Dore, R. A., Golinkoff, R. M., & Hirsh-Pasek, K. (2020). Children and screens. *Annual Review of Developmental Psychology*, *2*(3), article 1–3.24. https://doi.org/10.1146/annurev-devpsych-060320-095612

Heckhausen, H., & Kuhl, J. (1985). From wishes to action: The dead ends and short cuts on the long way to action. In M. Freese & J. Sabini (Eds.), *Goal directed behavior* (pp. 134–159). Routledge.

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*(4), 712–733. https://doi.org/10.1093/joc/jqy026

Jia, H., Wu, M., & Sundar, S. S. (2022). Do we blame it on the machine? In *Proceedings of the 55th Hawaii International Conference on System Sciences* (HICSS '22; pp. 388–397). https://doi.org/10.24251/HICSS.2022.047

Keijsers, M., & Bartneck, C. (2018). Mindless robots get bullied. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18; pp. 205–214). https://doi.org/10.1145/3171221.3171266

Kim, T., & Song, H. (2021). How should intelligent agents apologize to restore trust? Interaction effects between anthropomorphism and apology attribution on trust repair. *Telematics and Informatics*, *61*, article 101595. https://doi.org/10.1016/j.tele.2021.101595

Kretzschmar, K., Tyroll, H., Pavarini, G., Manzini, A., Singh, I., & NeurOx Young People's Advisory Group. (2019). Can your phone be your therapist? Young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support. *Biomedical Informatics Insights*, *11*, article 1178222619829083. https://doi.org/10.1177/1178222619829083

Lee, S. K., Kavya, P., & Lasser, S. C. (2021). Social interactions and relationships with an intelligent virtual agent. *International Journal of Human-Computer Studies*, *150*, article 102608. https://doi.org/10.1016/j.ijhcs.2021.102608

Lee, Y.-C., Yamashita, N., Huang, Y., & Fu, W. (2020). "I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. In *Proceedings of SIGCHI '20 Human Factors in Computing Systems* (CHI '20; pp. 1–12). https://doi.org/10.1145/3313831.3376175

Liu, B. (2018). Social presence in human-AI interaction: A proposal for non-anthropocentric social presence. In *Proceedings of the 18th Conference of the International Society for Presence Research (ISPR)*. ISBN: 978-0-9792217-6-7.

Liu, B. (2021). In AI we trust? Effects of agency locus and transparency on uncertainty reduction in human–AI interaction. *Journal of Computer-Mediated Communication*, *26*(6), 384–402. https://doi.org/10.1093/jcmc/zmab013

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, *21*(10), 625–636. https://doi.org/10.1089/cyber.2018.0110

Liu, B., & Wei, L. (2021). Machine gaze in online behavorial targeting: The effects of algorithmic "humanlikeness" on social presence and social influence. *Computers in Human Behavior*, *124*, article 106926. https://doi.org/10.1016/j.chb.2021.106926

Meng, J., & Dai, Y. N. (2021). Emotional support from AI Chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, *26*(4), 207–222. https://doi.org/10.1093/jcmc/zmab005

Meyer, J. R. (2009). Effect of primary goal on secondary goal importance and message plan acceptability. *Communication Studies*, *60*(5), 509–525. https://doi.org/10.1080/10510970903260343

Meyer, J. R. (2021, August 31). *Goals, plans, and action models*. Oxford Research Encyclopedia of Communication. https://doi.org/10.1093/acrefore/9780190228613.013.760

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

Muresan A., & Pohl, H. (2019). Chats with bots: Balancing imitation and engagement. In *Proceedings of SIGCHI '19 Human Factors in Computing Systems* (CHI '19; LBW0225). https://doi.org/10.1145/3290607.3313084

Pickard, M. D., & Roster, C. A. (2020). Using computer automated systems to conduct personal interviews: Does the mere presence of a human face inhibit disclosure? *Computers in Human Behavior*, *105*, 106197. https://doi.org/10.1016/j.chb.2019.106197

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

Rodríguez-Hidalgo, C. (2020). Me and my robot smiled at one another: The process of socially enacted communicative affordance in human-machine communication. *Human-Machine Communication*, *1*, 55–69. https://doi.org/10.30658/hmc.1.4

Roloff, M. E., & Janiszewski, C. A. (1989). Overcoming obstacles to interpersonal compliance a principle of message construction. *Human Communication Research*, *16*(1), 33–61. https://doi.org/10.1111/j.1468-2958.1989.tb00204.x

Schrader, D. C., & Dillard, J. P. (1998). Goal structures and interpersonal influence. *Communication Studies*, *49*, 276–293. https://doi.org/10.1080/10510979809368538

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

Sears, R. R. (1957). Identification as a form of behavioral development. In D. B. Harris (Ed.), *The concept of development* (pp. 149–161). University of Minnesota Press.

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

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2021). My chatbot companion—A study of human-chatbot relationships. *International Journal of Human-Computer Studies*, *149*, article 102601. https://doi.org/10.1016/j.ijhcs.2021.102601

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*(3), 272–280. https://doi.org/10.1080/08824096.2014.924337

Strait, M. K., Aguillon, C., Contreras, V., & Garcia, N. (2017). The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "Technology Takeover", and the unabashed sexualization of female-gendered robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (pp. 1418–1423). https://doi.org/10.1109/ROMAN.2017.8172490

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

Sundar, S. S., Jia, H., Waddell, T. F., & Huang, Y. (2015). Toward a theory of interactive media effects (TIME): Four models for explaining how interface features affect user psychology. In S. S. Sundar (Ed.), *The handbook of the psychology of communication technology* (pp. 47–86). Wiley-Blackwell. https://dx.doi.org/10.1002/9781118426456.ch3

Sundar, S. S., & Kim, J. (2019). Machine heuristic: When we trust computers more than humans with our personal information. In *Proceedings of SIGCHI '19 Human Factors in Computing Systems* (CHI '19; pp. 1–9). https://doi.org/10.1145/3290605.3300768

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

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., DeCero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*, *22*(3), article 16235. https://www.jmir.org/2020/3/e16235/

van Straten, C. L., Peter, J., Kühne, R., & Barco, A. (2021). The wizard and I: How transparent teleoperation and self-description (do not) affect children's robot perceptions and child-robot relationship formation. *AI & SOCIETY*, 1–17. https://doi.org/10.1007/s00146-021-01202-3

van Straten, C. L., Peter, J., Kühne, R., & Barco, A. (2022). On sharing and caring: Investigating the effects of a robot's self-disclosure and question-asking on children's robot perceptions and child-robot relationship formation. *Computers in Human Behavior*, *129*, article 107135. https://doi.org/10.1016/j.chb.2021.107135

Veletsianos, G., Scharber, C., & Doering, A. (2008). When sex, drugs, and violence enter the classroom: Conversations between adolescents and a female pedagogical agent. *Interacting with Computers*, *20*(3), 292–301. https://doi.org/10.1016/j.intcom.2008.02.007

Velez, J. A., Loof, T., Smith, C. A., Jordan, J. M., Villarreal, J. A., & Ewoldsen, D. R. (2019). Switching schemas: Do effects of mindless interactions with agents carry over to humans and vice versa? *Journal of Computer-Mediated Communication*, *24*(6), 335–352. https://doi.org/10.1093/jcmc/zmz016

Velez J. A., Schmierbach, M., Sherrick, B., Chae, S. W., Tan, R. R., & Rosenberg, K. A. (2021). A recipe for success: The effect of dyadic communication and cooperative gameplay on subsequent non-gaming task performance. *Human Communication Research*, *47*(4), 364–386. https://doi.org/10.1093/hcr/hqab008

Vitale, J., Tonkin, M., Herse, S., Ojha, S., Clark, J., Williams, M. A., Wang, X., & Judge, W. (2018, February). Be more transparent and users will like you: A robot privacy and user experience design experiment. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '18; pp. 379–387). https://doi.org/10.1145/3171221.3171269

Walther, J. B. (1992). Interpersonal effects in computer-mediated interaction. A relational perspective. *Communication Research*, *19*(1), 52–90. https://doi.org/10.1177/009365092019001003

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

Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-it, I-thou, I-robot: The perceived humanness of AI in human-machine communication. *Communication Studies*, *71*(3), 393–408. https://doi.org/10.1080/10510974.2020.1749683

Wilson, S. R., Aleman, C. G., & Leatham, G. B. (1998). Identity implications of influence goals: A revised analysis of face-threatening acts and application to seeking compliance with same-sex friends. *Human Communication Research*, *25*, 64–96. https://doi.org/10.1111/j.1468-2958.1998.tb00437.x

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

Zellou, G., Cohn, M., & Kline, T. (2021). The influence of conversational role on phonetic alignment toward voice-AI and human interlocutors. *Language, Cognition and Neuroscience*, *36*, 1298–1312. https://doi.org/10.1080/23273798.2021.1931372

Sex With Robots and Human-Machine Sexualities: Encounters Between Human-Machine Communication and Sexuality Studies

Marco Dehnert[1]

1 The Hugh Downs School of Human Communication, Arizona State University, Tempe, AZ, USA

Abstract

Sex robots are a controversial topic. Understood as artificial-intelligence enhanced humanoid robots designed for use in partnered and solo sex, sex robots offer ample opportunities for theorizing from a Human-Machine Communication (HMC) perspective. This comparative literature review conjoins the seemingly disconnected literatures of HMC and sexuality studies (SeS) to explore questions surrounding intimacy, love, desire, sex, and sexuality among humans and machines. In particular, I argue for understanding human-machine sexualities as communicative sexuotechnical-assemblages, extending previous efforts in both HMC and SeS for more-than-human, ecological, and more fluid approaches to humans and machines, as well as to sex and sexuality. This essay continues and expands the critical turn in HMC by engaging in an interdisciplinary exercise with theoretical, design, and use/effect implications in the context of sex robots.

Keywords: human-machine communication, sexuality studies, sex robot, assemblage, more-than-human

> *Contemporary academic discourse needs to move away from the idea of sexuality as a subject position, nicely and relatively stably wrapped under the epidermal cover of an individual human body, and develop instead a vocabulary about affective intensity, flux, and the sensual assembling of human and nonhuman elements into a pleasure machine.*—Lambevski, 2004, p. 305

CONTACT   Marco Dehnert ⓘ • The Hugh Downs School of Human Communication • Arizona State University • P.O. Box 871205, Tempe, AZ 85287-1205, USA • marco.dehnert@asu.edu

For an introductory price of less than $10,000, interested customers may purchase "affordable" sex robots from Abyss Creations' RealDoll, one of the leading manufacturers of sex robots. At the time of this writing, most sex robots consist of high-end sex dolls equipped with an artificial intelligence (AI)-enhanced robotic head. Fully robotic sex bots are nonetheless in the works by many companies worldwide. The case of sex robots opens particularly interesting questions that reverberate across many domains of society, from companionship and intimacy to therapeutic usage and questions regarding the (il)legality of child sex dolls (e.g., Chatterjee, 2020). Opinions range from calls for abolishing sex robots (Campaign Against Porn Robots, n.d.[1]; Richardson, 2016a, 2016b) to heralding the many social and individual benefits (Levy, 2007a), with a broad range of opinions located somewhere between these extremes (e.g., Ess, 2018).

In this essay, I explore human-machine sexualities at the intersection of human-machine communication (HMC) and sexuality studies (SeS). My aim in this essay is to engage in a comparative literature review that seeks to elaborate on the interdisciplinary intersections between the work in these two different fields, demonstrating how and why research on the subject of sex robots can inform work in HMC and how efforts in HMC can provide new insights for the study of sex robots, particularly from an SeS perspective. In doing so, I respond to Döring et al.'s (2020) call to increase the degree of theoretical elaboration of human-sex robot relations. With the arrival of interactive and communicative sex robots, I ponder the question, How can the bodies of literature in HMC and SeS enrich each other in the context of sex robots? In particular, by drawing on SeS in conjunction with HMC, I ask: In what ways do human-machine sexual relations alter our understanding of sexuality? What happens to our understanding of love and eroticism, intimacy and sexual closeness when the other is AI? In what ways does humans' interaction with sex bots affect ontologizing processes, or the drawing of boundaries between humans and machines (Fortunati & Edwards, 2021)? In conjoining the emerging field of HMC with the rich, critical, and incoherent body of SeS, I punctuate how machines reconstitute sexualities and work the fuzzy edges in response to Hearn's (2018) question, "what are the boundaries around [human] sexuality?" (p. 1368). More directly, what exactly constitutes the boundaries of human sexuality if the sexual partner is nonhuman?

This work continues earlier calls for mobilizing a critical perspective in HMC. Particularly in the realm of human-machine encounters where humans interact with machine-others in the creation of meaning (Fortunati & Edwards, 2020; Guzman, 2018), a critical communication perspective attunes us to the ways in which sociohistorical and cultural systems shape the ways in which humans make sense of machines (Dehnert & Leach, 2021). By turning to the rich context of human-machine sexualities and sex robots, I seek to further flesh out what a critical communication perspective to the study of machines and their co-creation of meaning with humans entails. In doing so, I argue that interdisciplinary approaches are necessary to fully capture the societal implications of machines, which I

---

1. In July of 2021, the organization formerly known as *Campaign Against Sex Robots* announced its name change to *Campaign Against Porn Robots* to reflect that, according to the campaign organizers, sexual activities involving robots are not "actual" sex but rather reflect processes of pornification and objectification (Campaign Against Porn Robots, n.d.; see also Danaher et al., 2017). The distinction between porn and sex related to robots seems to indicate the safeguarding of (human-to-human) notions of sex understood in the context of authenticity, intimacy, love, and connection, which are bypassed by the more-than-human framework of communicative sexuotechnical-assemblage put forth in this essay.

demonstrate by bringing SeS in conversation with HMC. HMC offers a rich contextual framework for making sense of human-machine sexualities, and sex robots constitute an intriguing context for investigating the boundaries of machines as communicative others. By conjoining HMC and SeS, I investigate where one draws the boundaries between sex robots, sex toys, and other emerging technologies in the broader realm of the sexual and between (sex) robots and (a/sexual) humans.

As such, the main goal of this essay is to argue that the case of sex robots illustrates the necessity for critical approaches to human-machine relations (in HMC) and to sexuality (in SeS) writ large. I begin this argument by reviewing the ways in which machines are cast as communicative others and further outline the implications of a critical communication perspective to HMC. After I tentatively differentiate sex robots from other technologies, I specifically utilize interdisciplinary more-than-human approaches to both machines and sexuality, extending Flore and Pienaar's (2020) notion of sexuotechnical-assemblage to describe the distinctly technological dimension of sexuality in human-machine relations. Further, by recasting sexuality as assemblage, I follow Fox and Alldred's (2013) approach which "shifts the location of sexuality away from bodies and individuals, toward the affective flow within assemblages of bodies, things, ideas, and social institutions, and the (sexual) capacities produced in bodies by this flow" (p. 770). Conjoining this work, and thereby extending Martinez's (2011) attention to *communicative* sexualities, I argue for describing human-machine sexualities as communicative sexuotechnical-assemblages. Finally, recasting sexuality in these geographic registers of relationality and assemblage responds to ongoing critiques of sexual science's continued exclusion of and violence toward sexual others via pathologization (Flore, 2014), compounded colonization (Balestrery, 2012), and normalization (Irvine, 1990; Somerville, 1994). I conclude this essay with theoretical and design implications.

HMC and the Machine-Other

HMC constitutes a rapidly growing field within the broader realm of communication studies focused on the ways in which humans interact with machine-others. What sets HMC apart from related fields is the focus on the communication processes between humans and machines in which the machine is not rendered as a channel through which humans communicate, but as "a communicative subject" with whom humans interact (Guzman, 2018, p. 12; Fortunati & Edwards, 2020). I use "machine-others" (rather than "machine" itself) to highlight this communicative subjectivization of the machine in HMC encounters. Understood as "*the creation of meaning among humans and machines*" (Guzman, 2018, p. 1, emphasis in original), HMC addresses topics such as agency (Banks & de Graaf, 2020), ontological boundaries (Edwards, 2018; Guzman, 2020), and the role and applicability of human-to-human scripts to human-machine encounters (Dehnert & Leach, 2021; Westerman et al., 2020).

In the context of HMC, the Computers Are Social Actors (CASA) paradigm—and, more recently, the Media Are Social Actors (MASA) paradigm (Lombard & Xu, 2021)—and constructivist approaches have been utilized to describe the ways in which humans relate to and interact with machines (e.g., Gambino et al., 2020; Nass & Moon, 2000; Westerman et al., 2020). Roughly (for reviews see Fortunati & Edwards, 2021; Westerman et al., 2020),

CASA/MASA and constructivist approaches explain the ways in which humans apply previously learned communication scripts to their encounters with machines. Conceiving of human-machine encounters through a Buberian I-Thou framework allows for the application of human-human communication theories to HMC contexts (Westerman et al., 2020). Recently, however, scholars have called for mobilizing a critical communication perspective in HMC to reconsider the ways in which the machine-other is otherized in human-machine encounters, asking questions related to ableism, gendered and sexed dynamics, as well as processes of racialization (Fortunati & Edwards, 2021; e.g., Davis & Stanovsek, 2021; Dehnert & Leach, 2021; Liu, 2021; Moran, 2020).

The context of sex robots offers unique vantage points for furthering critical perspectives in HMC. On the one hand, this context allows for drawing on the rich literature in SeS—where sexuality emerges as a sociohistorical formation, "a vital means of pleasure, interpersonal connection, personal efficacy, and acceptance of one's body and of self more generally" (Wilkerson, 2011, p. 194). Thus, in addition to previous critical work in HMC, the context of sex robots invites other intersectional markers of difference theorized in SeS, queer and trans (of color) criticism, and feminist and crip theories: In addition to dis/ability, this involves sex, sexuality, gender, race, class, and age, among others.

On the other hand, sex robots allow for exploring the complex implications of communicative sexual machine-others on intimate relations, including major social, legal, political, and ethical implications regarding the role of sex/uality for humans. After all, as Flore (2014) insists, "The birth of the sexual sciences, and the development of sexology and psychiatry, were and remain an attempt to define and delimit the meaning of being *human* itself" (p. 18, emphasis in original; see also Foucault, 1978). Similarly, HMC engages ontological questions regarding the divides between humans, machines, and animals (Edwards, 2018; Guzman, 2020). For instance, although CASA states that humans treat machines as if they were people, "we may not always respond to people in a very interpersonal way," meaning that human-human interaction is oftentimes heavily scripted (Westerman et al., 2020, p. 403). This opens profound questions about what exactly characterizes human-human relationality and how it differs from human-machine relationality (Dehnert, 2021; Westerman et al., 2020).

In the context of asexuality studies, Flore (2014) argued that "to be human is to be sexual" and outlined the ways in which sexuality is compulsory in the context of the human (p. 19). In the following section, I take up Flore's and other SeS scholars' insights regarding the ways in which sexual science and sexology pathologize, otherize, and violently exclude deviance in sexual behavior, orientation, and identity to add to the ongoing conversations on sex robots and their implications. I review existing research on sex robots from an interdisciplinary perspective to highlight the ripe potential of conjoining this body of literature with HMC, specifically punctuating and extending Flore and Pienaar's (2020) notion of sexuotechnical-assemblage.

Sex Robots: A Controversial Technology and/in HMC

Sex robots are a controversial and highly debated topic in lay and academic circles, including but not limited to legal, ethical, design, feminist, clinical, therapeutic, and other perspectives

(for a review, see Döring et al., 2020). After conducting their review of interdisciplinary literature focused on sex dolls and sex robots, Döring et al. (2020) conclude that

> Predictions of small and/or ambivalent effects might be more realistic [than the dystopian or utopian visions so prevalent in the literature] but are seldom discussed in the academic literature thus far, which seems to mirror some of the hype and scandalization observable in public media discourses. (p. 21)

As one potential explanation for these dramatized perspectives, Döring and colleagues name the fact that many people do not have firsthand experiences with sex robots due to their scarcity and cost. Some scholars project, however, that people will gain significantly more experience with sex bots and, in 2050, it will be not only common for folks to experience sex and love with machines (Levy, 2007a), but that humans might actually have *more* sex with robots than with other humans (Pearson, 2015; see Hauskeller, 2017, for an important critique of such transhumanist visions). In light of these rather bold projections, it is necessary to further our understanding of human-machine sexualities.

Technology and sexuality have been embraced by many scholars, given the myriad ways in which technology—broadly understood as biomedical, biomechanical, and biodigital—is related to sexuality. Scholars have written on technologies in the realm of the sexual, such as pharmaceuticals (Flore, 2018), technology in pornography (Dekker et al., 2021), and sex robots. In the literature on sex technologies, and also in public perception, sex toys emerge as one of the more prominent technologies. Understood as "material objects selected, created, and used to generate or enhance sexual arousal and pleasure in both solo and partnered sex" (Döring, 2021b, p. 1), sex toys include both commercially produced and homemade objects such as vibrators or masturbators. Recent models including wireless sex toys or otherwise digitized and connected sex toys have been the focal point of scholars and designers, usually labeled teledildonics (Flore & Pienaar, 2020). These technologies are marketed as sensory devices that allow for haptic or kinaesthetic interaction between partners across distance, or with the technology that responds to movement and touch. Typically paired with smartphone applications (apps), these devices allow for personalization and recording of personal preferences, promising "to increase sexual performance and pleasure through the algorithmic analysis of data" (Flore & Pienaar, 2020, p. 280). However, critics note that teledildonics reintroduce well-discussed issues regarding sexual safety and normativity. For instance, Sparrow and Karas (2020) argue that teledildonics allow for "rape by deception," or the risk of being deceived about the sexual partner's features and/or about which person the user was having "sex" with. Thus, while teledildonics promise increased sexual pleasure and intimacy, these connected technologies raise intriguing questions about intimacy, sexual practices, and human interaction.

But what distinguishes a sex robot from a sex doll, a teledildonics device, or any other sex toy? Cognizant of Fortunati and Edwards's (2021) insight that "Robots have such a multiform and mutant body that it becomes difficult to talk of robots' identity as well as of robots' capabilities" in a general sense (p. 16), and aware of the rapid technological advances in the fields of robotics and AI, any attempt at defining sex robots must inevitably be incomplete and tentative. Moreover, definitions and designs vary, particularly regarding the role and prominence of AI and other robotic features of sex robots. This also includes

differences in the level of sophistication as it relates to the AI personalities being sold to customers across various models. Current technology is rather rudimentary (Döring et al., 2020); however, sex robots currently on the market should not be confused with advanced machines depicted in science fiction, which are oftentimes imagined having sentience, consciousness, free will, and the like. In a recent attempt at defining sex robots, Döring (2021a) offers "human-like, full-body, anatomically correct, humanoid service robots of different materials, technologies, and price ranges that are designed and used to generate or enhance sexual arousal and pleasure in both solo and partnered sex" (p. 1). What sets sex robots apart from sex dolls, then, is that they are "equipped with sensors, actuators, and artificial intelligence" (p. 1)—in short, some sort of automated or mechanical technological features that allow the sex robot to move, talk, or otherwise interact with the human user.

In the case of the aforementioned Abyss Creations' RealDoll, for instance, customers can purchase AI-enhanced, robotic heads that can be added to sex dolls. Users can customize their sex dolls/robots in many ways, including body shape, skin tone, eye color, make-up, face, hair, and more. The AI-enhanced head allows for users to engage with their sex robot in various ways, including conversation, and the robotic head includes features such as eye movement, facial expression, as well as neck and mouth movement. With the accompanying app, users can fully customize their sex robot's personality, allowing the AI to learn the user's interests and preferences. RealDoll's dolls and robots are available as male, female, and transsexual models. Users can even purchase Bluetooth-enabled haptic vaginal sensors for the dolls which "can detect touch, movement, and transitions from mild arousal to orgasm" (RealDoll, n.d.a). In short, most contemporary sex robots are sex dolls enhanced with very limited AI and robotic features, oftentimes limited to specific body parts (head, vaginal sensors), and with limited interactive capabilities (conversation, eye movement, haptic feedback). Nonetheless, compared to sex dolls, these more interactive technological capabilities of sex robots are imagined leading to potentially rich relationships and shared, communicatively constructed meaning, where expected capabilities of sex robots involve hearing, recognizing objects, talking, or even taking initiative, among others (Scheutz & Arnold, 2016). The current sex robot market is advancing rapidly—as is the technology—but difficult to review, not least given the vast social stigma associated with this technology and its users. It is important to note that relations with each product differ, given variance in robotic capabilities, AI affordances, levels of sophistication, and user characteristics.

Notwithstanding these rather limited robotic and interactive functions of contemporary sex robots, scholars have expressed a variety of concerns and hopes in relation to sex robots (for reviews, see Döring et al., 2020; González-González et al., 2021). Although sparked by transhumanist researchers like Levy (2007a), the debate on sex robots has broadly considered domestic, commercial, and therapeutic use of sex robots (Döring et al., 2020). As such, target audiences and potential uses of sex robots vary, which shapes the production and design of sex robots (see Harper & Lievesley, 2020). These debates are additionally complicated by the "purely speculative" nature of claims about current and future effects as well as potential benefits and harms of sex robots and the scarcity of empirical studies thus far (Döring et al., 2020, p. 2). In addition to several edited collections (e.g., Bendel, 2020; Cheok et al., 2016; Cheok & Zhang, 2019; Danaher & McArthur, 2017; Zhou & Fischer, 2019) and monographs (e.g., Levy, 2007a), scholars have written about sex robots

related to their conceptualization and theory (e.g., Danaher, 2017a); engaged in legal and ethical considerations related to rape (e.g., Sparrow, 2017) or child robots (e.g., Chatterjee, 2020); investigated humans' perceptions of and attitudes toward sex robots (e.g., Middleweek, 2021; Scheutz & Arnold, 2016, 2017); examined potential therapeutic use of sex robots including health implications (Cox-George & Bewley, 2018) or potential use of sex robots for persons with disabilities (e.g., Fosch-Villaronga & Poulsen, 2021); critiqued sex robot representation in art and media (for a review see Döring & Poeschl, 2019); and have considered design questions (e.g., Danaher, 2019a).

Interestingly, Döring et al. (2020) did not report studies on sex robots conducted from a communication studies perspective, highlighting the need for communication scholars to contribute to this broad context and diverse literature. That is, with their focus on how meaning is created in human-machine interactions, HMC scholars can provide unique and novel insights into the characterization and understanding of sex robots. Thus, theorizing within HMC suggests an alternative perspective to what sets sex robots apart from adjacent technologies such as sex dolls, teledildonics, and other sex toys. In fact, as Döring et al. (2020) report, the current sex robot literature "often falls back on binary thinking" when it comes to conceptualizing sex robots:

> It categorizes the current sex robot as an inanimate object and mere masturbation aid without any sociability and is only willing to ascribe sociability to future imagined sex robots that are advanced to the point of indistinguishability from humans. Hence, the literature on sex robots often misses the key point that robots are more than mere masturbation aids due to anthropomorphization and that they are meaningful and possibly helpful precisely because they are not substitutes for real humans but are sociotechnical entities for parasocial use and play. (p. 20)

HMC has a lot to contribute in response to this diagnosis, given the ongoing theorizing of human-machine relationships with a focus on meaning-making in the field. Moreover, ongoing scholarly efforts in the CASA/MASA paradigms as well as ontologizing efforts (Edwards, 2018; Guzman, 2020) within HMC offer ample opportunities to contribute to research and theory of sex robots. In fact, recent efforts for a critical turn in HMC (Fortunati & Edwards, 2021) and posthuman perspectives (e.g., Betlemidze, 2022; Dehnert, 2021) provide useful theoretical backdrops for exploring sex robots and how human-machine sexualities alter our understanding of humans, machines, and sexualities. In the next section, I elaborate how these perspectives support understanding human-machine sexualities as communicative sexuotechnical-assemblages.

Human-Machine Sexualities as Communicative Sexuotechnical-Assemblages

This essay is not the first attempt at connecting more-than-human thought and other close relatives, such as posthumanism, new materialism, vital materialism, or object-oriented ontology to human-machine interactions (e.g., Betlemidze, 2022; Dehnert, 2021; Dehnert & Leach, 2021; Kubes, 2019; Ornella, 2009). For instance, in her critique of Richardson's

absolute stance against sex robots (Campaign Against Porn Robots, n.d.), which is built around a normative conception of "real sex," Kubes (2019) highlights the hidden, normative assumptions regarding "proper sex" and "proper love," and asks profound questions: "Does loving and feeling loved necessarily require its object to 'love back'? Or does it suffice, when the loving person *assumes* that their love is shared? I am leaning toward the latter" (Kubes, 2019, p. 4, emphasis in original). More directly located within a philosophical and ethical approach to HMC, scholars have described what Gerdes (2015) refers to as the social relational turn, specifically as it relates to the moral consideration of robots. Authors such as Coeckelbergh (2010, 2012), Gunkel (2012, 2018), and Gerdes have engaged in an ongoing conversation related to the moral standing of robots, with Coeckelbergh (2010, 2012) and Gunkel (2012, 2018) arguing more strongly for a social relational approach where moral status is not dependent on an entity's properties but viewed as socially constructed in the situated relationship, and Gerdes arguing for a human-centered framework. As Coeckelbergh (2010) argues, for a social relational approach to robot ethics, "moral significance resides neither in the object nor in the subject, but in the relation between the two" (p. 214). Although the question of moral consideration in the case of sex robots is important, as Ess (2016, 2018) demonstrates, these philosophical issues go beyond the scope of this essay. Nonetheless, in asking these kinds of questions, scholarly debates surrounding social relations, new materialism, and other more-than-human endeavors offer profound challenges to fundamental concepts in *both* SeS *and* HMC, questioning the concepts of subject, object, their relationship, their respective agency, and more.

Elsewhere (Dehnert, 2021), I have already engaged in a speculative exercise in what I call machine geographies—more-than-human communication geographies of human-machine encounters. In addition to the philosophical efforts related to the social relational turn described above, I employed geographical registers of agency, aesthetics, and ecology to outline what more-than-human approaches to HMC can look like. This includes, perhaps most profoundly, a recasting of agency in human-machine interactions as "relational, assemblage, fluid, in-between actors, as making-with, as achievement within networks, and becoming," thereby bypassing any considerations of communicative subject and object (Dehnert, 2021, p. 1154). Resonating with the aforementioned social relational turn, a more-than-human perspective allows for recasting human-machine interaction as relation. By focusing on relationality and the entanglements of humans and machines, scholars are not occupied with drawing fixed boundaries between humans and machines or with determining subject- and object-status in communicative encounters, but can embrace a "shift in focus from epistemological questions—such as what the objects 'represent'—to ontological questions about the kinds of qualities that they help to materialise or enact" (Flore & Pienaar, 2020, p. 283).

SeS scholars have also called for a similar shift in their respective field to move understandings of sexuality away from person-based definitions. The fields of sexual science and sexology are continuously critiqued by the more humanities- and critical-leaning SeS for medicalizing and "healthicizing" sexuality (notions such as "healthy sex drive" or "healthy sex behavior"), pathologizing and erasing non-normative sexual behaviors and identities, and for the continued ignorance toward the whiteness and racialized cisheterosexism so prominent among social scientific approaches to sex and sexuality (e.g., Balestrery, 2012; Flore, 2014; Irvine, 1990; Manalansan, 2013; Somerville, 1994). Work such as Balestrery's

explication of compounded colonization highlight the mutually informing and enforcing, racialized and sexualized ideological paradigms in sexology and sexual science. Efforts by Marxist feminists showcase the intricated connections between market-driven and labor-related notions of sex as a transaction in a neoliberal context, particularly as it relates to consumption (e.g., Miller-Young, 2014; Zatz, 1997). Similarly, studies of racialized sexualities (cf. Ferguson, 2007), specifically Black sexualities of anti-respectability in the context of sex work among Black queer women femmes (Glover & Glover, 2019) or Black women in pornography (Miller-Young, 2014) challenge not only the alive-and-well scientific racism and pathologizing of non-normative sexual subjectivities, practices, and identities, but also call for theorizing from the perspective of those marginalized by normativities constituted around white, cishetero, abled, settler perspectives.

Together, these critical endeavors in SeS understand sexuality *not* as the "biological, psychological and social processes associated with sexual desire, sensation, arousal, attraction and pleasure" (Fox & Alldred, 2013, p. 785n1), but as a sociohistorical formation that is constructed, imbricated by stratified formations of power, and itself a stratifying force on a societal level (Foucault, 1978). Such a shift away from a person-centered approach to sexuality resonates with Martinez's (2011) musings on the communicative nature of sexuality, the study of which means "to locate the phenomenon of sexuality within the intricacies of our immediate and embodied interconnection with the social and cultural world in which we are situated" (p. 11). Thusly reframed, sexuality is no longer confined to the property of persons nor the "intimate" spaces between people, or around one person individually, but is conceptualized as a fundamental mechanism of and in societies—a social technology in the Foucauldian sense—that both disciplines bodies while opening up space for resistance. In this sense,

> sexual agency [is] not merely [understood] as the capacity to choose, engage in, or refuse sex acts, but as a more profound good that is in many ways socially based, involving not only a sense of oneself as a sexual being but also a larger social dimension in which others recognize and respect one's identity. (Wilkerson, 2011, p. 195)

Additionally, Martinez (2011) highlights that this revisited notion of sexuality "is actualized only by the virtue of communicative processes in which we are always and inescapably *situated*" (p. 11, emphasis in original). Resonating with a social relational turn in the moral consideration of robots (Coeckelbergh, 2012), then, these perspectives prioritize the situated relation in the description of sexuality over entities' ontological properties.

This challenge to person-based understandings of sexuality and communication, as well as sexual and communicative relations and agency, can be further complemented by Deleuzo-Guattarian (1988) perspectives on assemblage. Drawing from anti-essentialist, anti-humanist, and Deleuzo-Guattarian thought, Fox and Alldred (2013) offer sexuality-assemblage as a theoretical move that overturns anthropocentric specters of sexuality focused on the individual human body. In their thick, sociological rethinking of sexuality as assemblage, they shift "the location of sexuality . . . toward the affective flow within assemblages of bodies, things, ideas and social institutions, and the (sexual) capacities produced in bodies by this flow" (p. 770). This rethinking of sexuality as assemblage, in relational,

ecological, and interconnected ways, resonates strongly with the similar shift in HMC out-lined above. Not only do Fox and Alldred draw on similar theoretical bodies of thought, they also apply Deleuze and Guattari's notion of assemblages as desiring-machines to highlight the role of affective flows, processual interactions, and the dissolution of sexual subject-object pairings when shifting to sexuality-assemblage (e.g., Deleuze & Guattari, 1988). Unlike person-based notions of sexual agency, then, Fox and Alldred (2013) consider agency as the "capacity to affect or be affected" (p. 772) and dislocate any considerations regarding sexual object-choice or musings on the object of someone's desire by pointing out that "productive desire makes affect flow in assemblages" (Fox & Alldred, 2013, p. 773). Crucially, this shift toward assemblage allows Fox and Alldred to reconceptualize sexuality as "*the flow of affect in a sexuality-assemblage*," manifesting in two ways: First, sexuality as the "deterritorializing, nomadic and rhizomic flow of affect between and around bodies and other relations" as a sociohistorical formation that suffuses much if not all of social life, and second, in the form of individual sexual desire, as a "territorialization of an impersonal, non-human and nomadic sexuality" (pp. 767–777, emphasis in original). This approach allows for recasting anthropocentric, humanist idea(l)s of sexuality, sexual agency, desire, resistance, and the relevance of sexuality as "a fundamental experience of what it means to be human" (Ornella, 2009, p. 318; Flore, 2014).

In their more-than-human work on "data-driven intimacy" and teledildonics, Flore and Pienaar (2020) adapt Fox and Alldred's (2013) sexuality-assemblage and explicitly connect it to technology such as data, algorithms, and wireless sex toys. They offer nuance to Fox and Alldred's (2013) above conceptualization of sexuality-assemblage and outline what they term sexuotechnical-assemblage, "a term that points to the range of actors and relations imbricated in teledildonic sex [and in human-machine sexualities, I would add], including human bodies and desires, sexual practices, technological devices, internet connections, intimate data, and neoliberal understandings of sexual health" and normative sexual desire, practices, and identities (Flore & Pienaar, 2020, p. 285).

Together, machine geographies (Dehnert, 2021) and sexuotechnical-assemblage (Flore & Pienaar, 2020; Fox & Alldred, 2013) allow for reimagining human-machine sexualities in important and useful ways. First, assembled sexual and communicative (or sexuo-communicative) relationalities between humans and machines ultimately displace questions that seek to investigate the ways in which machines emerge as communicative and sexual subjects in human-machine encounters. This does *not* imply a recurse to casting machines as mere objects or channels of human-human interaction and desire, given that assemblage simultaneously displaces the object. In this way, directionality of both desire and communication in (sexual) human-machine encounters is bypassed in favor of entan-gled, relational, affective, and aesthetic flows.

Second, adopting the perspective of sexuotechnical-assemblage for human-machine encounters reconceptualizes ongoing scholarly concerns regarding machine agency by dis-placing humanist attempts at locating agency within the machine. For instance, in their rich treatise on the foundations of erobotics (the transdisciplinary field concerned with artificial erotic agents), Dubé and Anctil (2021) describe erobots as agents by using the "broadest definition recognized and commonly used" in AI, robotics, and machine learning, where "the agency of machines refers to their capability to act intelligently in and on the world to achieve objectives of their own" (p. 1207). Reconceptualizing agency as not inherent to a

machine agent also speaks to the aforementioned social relational turn in the moral consideration of robots. In this vein, communicative sexuotechnical-assemblage focuses on the affective flow between inter-actants in relational, ecological, and assembled relations.

Third, the assemblage-perspective offers generative, alternative conceptualizations in response to, arguably decidedly, anthropocentric entry points into philosophical, ethical, and feminist conversations related to sex robots. For instance, in his ethical musings on the possibility of sex, love, and intimacy with sex robot, Ess (2018) concludes that "sexbots, as zombies lacking first-person phenomenal consciousness, genuine emotions, and (embodied) desire, will only be able to fake emotions" (p. 253). For Ess (2016, 2018), then, due to these shortcomings, it will be impossible to reach "complete sex," a high ethical standard for sexual relationships which is characterized by mutual desire and respect. Albeit holding a more radical position, the arguments put forth by what Danaher (2019a) calls "anti-sexbot feminism" (e.g., Campaign Against Porn Robots, n.d.; Richardson, 2016a, 2016b) take a similar stand toward what qualifies as proper, good, or authentic sex, which is therefore only limited to human-human sexuality. In response to these arguments, Danaher (2019a) draws on sex-positive feminist perspectives to articulate how we might build better sex robots, rather than follow Richardson and others' call for restricting them (see also Danaher & McArthur, 2017; Danaher et al., 2017). Pointing out the necessity for feminist insights into the content, process, and context for their creation, Danaher's (2019a) work represents the potential for conjoining critical perspectives of sex robots with their production. Importantly, an assemblage-perspective as argued for in this essay does not sidestep these important conversations, which matter in the context of child sex robots, for instance (Danaher, 2019b). Rather, it allows for alternative entry points that seek to increase the degree of theoretical elaboration of human-sex robot relations (Döring et al., 2020).

And, finally, by displacing concerns related to subject-object divides as well as notions of individualized agency, the assemblage-perspective both implicates the sociocultural in the intimate, and the intimate in the sociocultural. Returning to Fox and Alldred's (2013) language of (de)territorializing sexuality, they recognize that, while affective flows of/in sexuality-assemblages are unrestricted, they are often highly limited ("territorialized") based on individual and sociocultural contexts: "Sexual attraction, sexual preferences and proclivities are . . . territorializations toward particular objects of desire, consequent upon the particular mix of relations and affects deriving from physical and social contexts, experience and culture" (p. 775). As such, an assemblage-perspective is neither naïve toward norms and normativities (or scripts, cf. Dehnert & Leach, 2021) nor forecloses resistance and a rescripting of these cultural norms; in fact, the territorializing, deterritorializing, and reterritorializing dynamics within the sexuality-assemblage allow for resisting, reshaping, and transforming compulsory forms of sexuality (Emens, 2014; Flore, 2014). Therefore, human-machine sexualities as assemblage offer ripe entry points for critical approaches, both from an HMC and an SeS perspective.

Implications for HMC and SeS

In this comparative literature review, I have conjoined two rather disconnected bodies of literature in an effort to revisit sex robots in/as human-machine sexualities. I argued that, in addition to sensory, robotic, and AI-components (Döring, 2021a), what sets sex

robots apart from other sexual technologies is not so much their status as agents in sexual human-machine encounters, but rather the ways in which humans may relate with them; said differently, their capacity to create meaning with humans in human-machine sexualities, or communicative sexuotechnical-assemblages. In doing so, I respond to not only more-than-human efforts in HMC, but also to calls for fluidifying academic discourse on sexuality (Lambevski, 2004).

Theoretical Implications

Theorizing human-machine sexualities as communicative sexuotechnical-assemblages by conjoining rather disconnected bodies of literature and theoretical perspectives responds to recent calls for transdisciplinary and interdisciplinary approaches to the study of sex robots (e.g., Döring et al., 2020; Zhou & Fischer, 2020). Collectively, these approaches allow for more nuanced perspectives of sex robots that reject both the utopian visions of unending pleasure with sex robots (Levy, 2007a; Ornella, 2009) and the dystopian fears of those who call for abolishing sex robots (Campaign Against Porn Robots, n.d.). Grounded in communication and sexuality studies, and adjacent fields, the concept of communicative sexuotechnical-assemblage adds to existing, more philosophical and ethical projects in the context of machines generally (Coeckelbergh, 2010, 2012; Gerdes, 2015; Gunkel, 2012, 2018) and of sex robots specifically (e.g., Danaher, 2017b; Ess, 2016, 2018). By examining the affective flows between humans and machines in sexual encounters, scholars can track the various ethical, legal, sociological, and communicative issues addressed in the literature and raised by sex robots. This also includes a move beyond the binary thinking in the current literature, identified by Döring et al. (2020) and already elaborated on above. In so doing, human-machine sexualities continue and extend the recent critical turn in HMC by specifically turning to sex and sexuality as ripe *contexts*, and SeS as rich *resources* for critical efforts in HMC.

For instance, drawing on critical perspectives challenges reductionistic and problematic conceptions of sex work in debates on sex robots (Kubes, 2019). Authors such as Richardson (2016a, 2016b) or Levy (2007b), among others, who compare robot sex with human sex work, tend to fall back on understandings of sex work that frame the sex worker as "objectified and instrumentalized" (Danaher, 2017b, pp. 110–111) or as "reduced to a *thing*" (Richardson, 2016b, p. 291, emphasis in original). Critical SeS and, in particular, Marxist feminist, Black queer, crip, trans, and trans of color approaches offer a dramatically different and resistant understanding of sex work, guided by anti-respectability politics that highlight how "community members thrive despite existing in a hostile world unconcerned with their survival" (Glover & Glover, 2019, p. 172). Moreover, Danaher (2014) offered a nuanced understanding of sex work as it relates to what he describes as technological unemployment, or the displacement of human sex workers by the advent of sophisticated sex robots. This effort continues specifically Marxist and materialist understandings that correctly frame sex work in the context of labor and market dynamics. Drawing on critical non-white, non-cisheterosexist, and non-cisheteropatriarchal accounts of sexuality therefore simultaneously resists reductionistic accounts of sex work and sexuality writ large and opens up different ways to theorize and practice sexuality—both among humans and between humans and nonhumans.

Another implication based on the critical sensibilities inherent to communicative sexuotechnical-assemblages in human-machine sexual encounters is the ongoing critique of transhumanist utopias and fantasies in relation to sex robots (e.g., Lakshmanan, 2021). Specifically from the perspective of (critical) disability studies, transhumanist desires to improve, enhance, and perfect the human body either by modifying human bodies or by replacing "deficient" and "defunct" human bodies with better, and more "perfect" machine-others must grapple with ableist discourse of "curing" and "overcoming" disability (Hauskeller, 2017). Similar dynamics between sexuality and dis/ability have been centered by scholars in a/sexuality studies (e.g., Flore, 2014; Kafer, 2013; Wilkerson, 2011); efforts that consistently critique the normative formations of a "sexual" body, a "healthy" body, and "healthy" sexuality writ large. Human-machine sexualities navigate these important critiques by disregarding any consideration of enhancing the sexual other or sexuality as such, given its focus on the flows of affect between entangled partners. Nonetheless, dynamics of enhancement of intimacy, pleasure, and desire must be thoroughly examined in the context of emerging sexual technologies, including sex robots, particularly as it relates to labor, reproduction, and dynamics of the neoliberal market (Atanasoski & Vora, 2020).

Furthermore, human-machine sexualities embrace more-than-human and geographical registers in both HMC (cf. Betlemidze, 2022; Dehnert, 2021) and SeS (Flore & Pienaar, 2020; Fox & Alldred, 2013) that fluidify (Lambevski, 2004) individualistic and humanistic understandings of the sexuo-communicative subject-object relation in human-machine sexualities. These perspectives raise questions such as, what exactly is meant by "communicative subject" in entangled and assembled human-machine interactions? Resonating with the social relational turn, this essay continues rethinking agency, interactivity, directionality of communication and desire, and more, in ecological terms of affective flows.

Additionally, this broadening of conceptualizing the (sexual) communication practices between humans and machines comes along with rethinking the role of sexuality for the human. When humans engage with sex robots in communicative sexuotechnical-assemblages via human-machine sexual encounters, "what are the boundaries around [*human*] sexuality" (Hearn, 2018, p. 1368)? Insights from asexuality studies scholars demonstrate that, through discourses, instruments, and institutions, "'sexuality' effectively became tied to humanity" (Flore, 2014, p. 18). As such, ongoing work by asexuality studies scholars works toward delinking the intimate, compulsory relationship between being human and being sexual by making space for alternative modes of being and doing. Offering an alternative to person-based understandings of sexuality and sexual practices, more-than-human and assemblage-approaches resist clear-cut boundaries of (human) sexuality, thereby embracing the messiness of sexuality (Manalansan, 2013), even (or particularly) in the "sterile" context of machines.

Design Implications

Authors have predominantly critiqued representation and design of sex robots in relation to sexualized and exaggerated images of the female body or engaged in speculative musing on the ethical design of future sex robots (Döring et al., 2020). There is an insignificant integration of academic research and the design of sex robots, prompting calls for an integrated understanding of sex robots that recognizes the sociotechnical development and nature

of sex robots (Danaher, 2019a). Relatedly, designers can draw from SeS by being clearer about the distinction between gender and sex in robot design. Recall that RealDoll (n.d.b) offers sex robots in "three different gender orientations: Male, female and transsexual" (p. 6). Clearly, these descriptors refer neither to *gender* nor to an *orientation*, which calls for feminist and other critical approaches to more accurately imitate sex and gender in sex robot design. Finally, shifting toward an assemblage-framework allows for broadening the design of sex robots writ large, where "the obvious question we have to ask is: why should a sex robot look like a human?" (Kubes, 2019, p. 10). Displacing concerns for human and nonhuman subjects and objects allows for broadening our understanding of what a sex robot can be and can look like. Modeling sex robots after different fantasies than the male gaze so prevalent in pornographic and other representational accounts is one way to navigate the expectations surrounding humanoid robots (Danaher, 2019a).

Implications for the Use and Effects of Sex Robots

Reviewers (Döring et al., 2020; González-González et al., 2021) have identified a significant lack of empirical research conducted on actual use patterns and contexts as well as user behavior, leaving most claims about effects squarely within philosophical, ethical, and speculative realms (see Harper & Lievesley, 2020). Nonetheless, the assemblage-perspective put forth in this essay allows at least for comments on rough implications regarding the therapeutic use of sex robots and child robots, one of the most controversial components of an already highly controversial topic (e.g., Chatterjee, 2020). Critical communication and SeS perspectives call for a nuanced understanding of "therapeutic," paying particular attention to undergirding systems of belief that target a/sexual, disabled, and non-normative others in particular ways (Kafer, 2013; Wilkerson, 2011). Conjoining disability, SeS, and queer perspectives, the therapeutic use of sex robots must always be understood in the context of larger systems of meaning—which is reflected in the territorializing-deterritorializing-reterritorializing dynamics of the sexuality-assemblage (Fox & Alldred, 2013). That is, any calls for therapeutic use of sex robots must be critiqued: Therapy for whom, why, in what ways, and based on what grounds? Similarly, the case of child sex robots must be evaluated in the sociocultural context (Danaher, 2019b). Clearly, these debates are far from settled and require thorough, interdisciplinary contributions from academics, designers/manufacturers, and the general public.

Conclusion

With most thought and reflections on sex robots being confined to speculation at this time, this essay serves as a contribution to the ongoing, important debates on sex robots by conjoining two seemingly disconnected bodies of literature—HMC and SeS. Ongoing interdisciplinary work is needed as scholars make sense of current and future technological advancements in the realm of the sexual. I have specifically called for and extended efforts in the critical theorizing of sex robots in particular, and machines writ large. Questions regarding intimacy, love, sex, and desire have occupied humans for thousands of years. Reconceptualizing sex robots in the realm of human-machine sexualities, or communicative sexuotechnical-assemblages, allows for addressing the ways in which affective flows

between humans and machines constitute sexuality, as well as recognizing the ways in which notions of sex and sexuality are always tied to larger, deterritorialized systems of meaning. Current and emerging technologies such as teledildonics, AI, and physical sex robots offer vibrant potential for sex and sexuality, lying somewhere between utopian hopes for orgasmic frenzies and dystopian fears of sterile and stale numbing down. The boundaries of (human) sexuality are broad, fringy, messy, and oftentimes unclear. This is even more so the case as technology, and in particular sex robots, become increasingly entangled in human sexual relations.

Author Biography

Marco Dehnert (MA, Arizona State University) is a doctoral student in the Hugh Downs School of Human Communication at Arizona State University. His research interests focus on critical/cultural approaches to communication and rhetorical studies. With a broad focus on how one recognizes and disavows the Other, his work is situated at the intersections of critical and intersectional rhetorics, queer and transgender communication studies, qualitative inquiry, and human-machine communication.

https://orcid.org/0000-0002-7456-0743

References

Atanasoski, N., & Vora, K. (2020). Why the sex robot becomes the killer robot: Reproduction, care, and the limits of refusal. *Spheres, 6.* https://web.archive.org/web/20200918225712/ https://spheres-journal.org/contribution/why-the-sex-robot-becomes-the-killer-robot-reproduction-care-and-the-limits-of-refusal

Balestrery, J. E. (2012). Intersecting discourses on race and sexuality: Compounded colonization among LGBTTQ American Indians/Alaskan Natives. *Journal of Homosexuality, 59*(5), 633–655. https://doi.org/10.1080/00918369.2012.673901

Banks, J., & de Graaf, M. M. A. (2020). Toward an agent-agnostic transmission model: Synthesizing anthropocentric and technocentric paradigms in communication. *Human-Machine Communication, 1*, 19–36. https://doi.org/10.30658/hmc.1.2

Bendel, O. (2020). (Ed.). *Maschinenliebe: Liebespuppen und Sexroboter aus technischer, psychologischer und philosophischer Perspektive*. Springer.

Betlemidze, M. (2021). Traversing anthropocentric horizons with Her: Trans-corporeal surrogacy, enchantment, and disenchantment in human-machine assemblage. *Journal of Communication Inquiry*, 46(2), 206–224. https://doi.org/10.1177/01968599211041107

Campaign Against Porn Robots. (n.d.). *Home.* https://web.archive.org/web/20211108183623/ https://campaignagainstsexrobots.org/

Chatterjee, B. B. (2020). Child sex dolls and robots: Challenging the boundaries of the child protection framework. *International Review of Law, Computers & Technology, 34*(1), 22–43. https://doi.org/10.1080/13600869.2019.1600870

Cheok, A. D., Devlin, K., & Levy, D. (2016). (Eds.). *Love and sex with robots: Second International Conference, LSR 2016, London, UK, December 19-20, 2016*. Springer.

Cheok, A. D., & Zhang, E. Y. (2019). *Human-robot intimate relationships*. Springer.

Coeckelbergh, M. (2010). Robot rights? Towards a social-relational justification of moral consideration. *Ethics and Information Technology, 12*, 209–221. https://doi.org/10.1007/s10676-010-9235-5

Coeckelbergh, M. (2012). *Growing moral relations: Critique of moral status ascription*. Palgrave Macmillan.

Cox-George, C., & Bewley, S. (2018). I, sex robot: The health implications of the sex robot industry. *BMJ Sexual & Reproductive Health, 44*, 161–164. http://dx.doi.org/10.1136/bmjsrh-2017-200012

Danaher, J. (2014). Sex work, technological unemployment and the basic income guarantee. *Journal of Evolution and Technology, 24*(1), 113–130. https://jeet.ieet.org/index.php/home/article/view/17

Danaher, J. (2017a). Should we be thinking about robot sex? In J. Danaher & N. McArthur (Eds.), *Robot sex* (pp. 3–14). MIT Press.

Danaher, J. (2017b). The symbolic-consequences argument in the sex robot debate. In J. Danaher & N. McArthur (Eds.), *Robot sex* (pp. 103–131). MIT Press.

Danaher, J. (2019a). Building better sex robots: Lessons from feminist pornography. In Y. Zhou & M. H. Fischer (Eds.), *AI love you: Developments in human-robot intimate relationships* (pp. 133–148). Springer.

Danaher, J. (2019b). Regulating child sex robots: Restriction or experimentation? *Medical Law Review, 27*(4), 553–575. https://doi.org/10.1093/medlaw/fwz002

Danaher, J., Earp, B., & Sandberg, A. (2017). Should we campaign against sex robots? In J. Danaher & N. McArthur (Eds.), *Robot sex* (pp. 47–71). MIT Press.

Danaher, J., & McArthur, N. (Eds.). (2017). *Robot sex*. MIT Press.

Davis, D. Z., & Stanovsek, S. (2021). The machine as an extension of the body: When identity, immersion, and interactive design serve as both resource and limitation for the disabled. *Human-Machine Communication, 2*, 121–135. https://doi.org/10.30658/hmc.2.6

Dehnert, M. (2021). Communication geographies of human-machine understanding: Entangled agencies, synthetic aesthetics, and machine matterings. *Communication Studies, 72*(6), 1146–1159. https://doi.org/10.1080/10510974.2021.2011360

Dehnert, M., & Leach, R. B. (2021). Becoming human? Ableism and control in *Detroit: Become Human* and the implications for human-machine communication. *Human-Machine Communication, 2*, 137–152. https://doi.org/10.30658/hmc.2.7

Dekker, A., Wenzlaff, F., Biedermann, S. V., Briken, O., & Fuss, J. (2021). VR porn as "empathy machine"? Perception of self and others in virtual reality pornography. *The Journal of Sex Research, 58*(3), 273–278. https://doi.org/10.1080/00224499.2020.1856316

Deleuze, G., & Guattari, F. (1988). *A thousand plateaus*. Athlone.

Döring, N. (2021a). Sex dolls and sex robots. In A. D. Lykins (Ed.), *Encyclopedia of sexuality and Gender*. Springer. https://doi.org/10.1007/978-3-319-59531-3_63-1

Döring, N. (2021b). Sex toys. In A. D. Lykins (Ed.), *Encyclopedia of sexuality and gender*. Springer. https://doi.org/10.1007/978-3-319-59531-3_62-1

Döring, N., Mohseni, R., & Walter, R. (2020). Design, use, and effects of sex dolls and sex robots: Scoping review. *Journal of Medical Internet Research, 22*(7), e18551. https://doi.org/10.2196/18551

Döring, N., & Poeschl, S. (2019). Love and sex with robots: A content analysis of media representations. *International Journal of Social Robotics, 11*, 665–677. https://doi.org/10.1007/s12369-019-00517-y

Dubé, S., & Anctil, D. (2021). Foundations of erobotics. *International Journal of Social Robotics, 13*, 1205–1233. https://doi.org/10.1007/s12369-020-00706-0

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

Emens, E. F. (2014). Compulsory sexuality. *Stanford Law Review, 66*, 303–386. https://dx.doi.org/10.2139/ssrn.2218783

Ess, C. M. (2016). What's love got to do with it? Robots, sexuality, and the arts of being human. In M. Nørskov (Ed.), *Social robots: Boundaries, potential, challenges* (pp. 57–79). Ashgate.

Ess, C. M. (2018). Ethics in HMC: Recent developments and case studies. In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 237–257). Peter Lang.

Ferguson, R. A. (2007). Relevance of race for the study of sexuality. In G. E. Haggerty & M. McGarry (Eds.), *Companion to lesbian gay bisexual transgender studies* (pp. 106–123). Blackwell.

Flore, J. (2014). Mismeasures of asexual desires. In K. J. Cernakowski & M. Milks (Eds.), *Asexualities: Feminist and queer perspectives* (pp. 17–34). Routledge.

Flore, J. (2018). Pharmaceutical intimacy: Managing female sexuality through Addyi. *Sexualities, 21*(4), 569–586. https://doi.org/10.1177/1363460717731933

Flore, J., & Pienaar, K. (2020). Data-driven intimacy: Emerging technologies in the (re)making of sexual subjects and 'healthy' sexuality. *Health Sociology Review, 29*(3), 279–293. https://doi.org/10.1080/14461242.2020.1803101

Fortunati, L., & Edwards, A. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication, 1*, 7–18. https://doi.org/10.30658/hmc.1.1

Fortunati, L., & Edwards, A. (2021). Moving ahead with human-machine communication. *Human-Machine Communication, 2*, 7–28. https://doi.org/10.30658/hmc.2.1

Fosch-Villaronga, E., & Poulsen, A. (2021). Sex robots in care: Setting the stage for a discussion on the potential use of sex robots for persons with disabilities. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction (HRI '21 Companion), March 8–11, 2021, Boulder, CO, USA*. ACM. https://doi.org/10.1145/3434074.3446907

Foucault, M. (1978). *The history of sexuality, volume 1: An introduction*. Pantheon.

Fox, N. J., & Alldred, P. (2013). The sexuality-assemblage: Desire, affect, anti-humanism. *The Sociological Review, 61*(4), 769–789. https://doi.org/10.1111/1467-954X.12075

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

Gerdes, A. (2015). The issue of moral consideration in robot ethics. *ACM SIGCAS Computers & Society, 45*(3), 274–280. https://doi.org/10.1145/2874239.2874278

Glover, S. T., & Glover, J. K. (2019). "She ate my ass and my pussy all night": Developing illicit eroticism, funk, and sex work among Black queer women femmes. *American Quarterly, 71*(1), 171–177. https://doi.org/10.1353/aq.2019.0010

González-González, C. S., Gil-Iranzo, R. M., & Paderewski-Rodríguez, P. (2021). Human-robot interaction and sexbots: A systematic literature review. *Sensors, 21*(216). https://doi.org/10.3390/s21010216

Gunkel, D. J. (2012). *The machine question: Critical questions on AI, robots, and ethics*. MIT Press.

Gunkel, D. J. (2018). *Robot rights*. MIT Press.

Guzman, A. L. (2018). Introduction: "What is human-machine communication, anyway?" In A. L. Guzman (Ed.), *Human-machine communication: Rethinking communication, technology, and ourselves* (pp. 1–28). Peter Lang.

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

Harper, C. A., & Lievesley, R. (2020). Sex doll ownership: An agenda for research. *Current Psychiatry Reports, 22*(54). https://doi.org/10.1007/s11920-020-01177-w

Hauskeller, M. (2017). Automatic sweethearts for transhumanists. In J. Danaher & N. McArthur (Eds.), *Robot sex* (pp. 203–218). MIT Press.

Hearn, J. (2018). Where are the boundaries of sexuality? Hovering in a zone of uncertainty between sexualities and non-sexualities. *Sexualities, 21*(8), 1368–1373. https://doi.org/10.1177/1363460718785108

Irvine, J. M. (1990). From difference to sameness: Gender ideology in sexual science. *The Journal of Sex Research, 27*(1), 7–24. https://doi.org/10.1080/00224499009551539

Kafer, A. (2013). *Feminist, queer, crip*. Indiana University Press.

Kubes, T. (2019). New materialist perspectives on sex robots: A feminist dystopia/utopia? *Social Sciences, 8*(8), 224. https://doi.org/10.3390/socsci8080224

Lakshmanan, N. (2021). The misogyny of transhumanism. In N. Rambukkana (Ed.), *Intersectional automations: Robotics, AI, algorithms and equity* (pp. 99–118). Lexington.

Lambevski, S. A. (2004). Movement and desire: On the need to fluidify academic discourse on sexuality. *GLQ: A Journal of Lesbian and Gay Studies, 10*(2), 304–308.

Levy, D. (2007a). *Love and sex with robots: The evolution of human-robot relationships*. Harper Collins.

Levy, D. (2007b). Robot prostitutes as alternatives to human sex workers. In *IEEE international conference on robotics and automation, Rome*. https://web.archive.org/web/20090306164319/http://www.roboethics.org/icra2007/contributions/LEVY%20Robot%20Prostitutes%20as%20Alternatives%20to%20Human%20Sex%20Workers.pdf

Liu, J. (2021). Social robots as the bride? Understanding the construction of gender in a Japanese social robot product. *Human-Machine Communication, 2*, 105–120. https://doi.org/10.30658/hmc.2.5

Lombard, M., & Xu, K. (2021). Social responses to media technologies in the 21st century: The media are social actors paradigm. *Human-Machine Communication, 2*, 29–55. https://doi.org/10.30658/hmc.2.2

Manalansan, M. F. (2013). Queer worldings: The messy art of being global in Manila and New York. *Antipode, 47*(3), 566–579. https://doi.org/10.1111/anti.12061

Martinez, J. M. (2011). *Communicative sexualities: A communicology of sexual experience*. Lexington.

Middleweek, B. (2021). Male homosocial bonds and perceptions of human-robot relationships in an online sex doll forum. *Sexualities, 24*(3), 370–387. https://doi.org/10.1177/1363460720932383

Miller-Young, M. (2014). *A taste for brown sugar: Black women in pornography*. Duke University Press.

Moran, T. C. (2020). Racial technological bias and the white, feminine voice of AI VAs. *Communication and Critical/Cultural Studies, 18*(1), 19–36. https://doi.org/10.1080/14791420.2020.1820059

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

Ornella, A. D. (2009). Posthuman pleasures: Transcending the human-machine boundary. *Theology & Sexuality, 15*(3), 311–328. https://doi.org/10.1558/tse.v15i3.311

Pearson, I. (2015). The future of sex report: The rise of the robosexuals. A Futurizon Report—Bondara. https://web.archive.org/web/20160310000549/http://graphics.bondara.com/Future_sex_report.pdf

RealDoll (n.d.a). Home. https://web.archive.org/web/20210801121434/https://www.realdoll.com/

RealDoll (n.d.b). *Media Kit* [PDF]. https://web.archive.org/web/20211228205509/https://www.realdoll.com/wp-content/uploads/2020/05/RealDoll-Abyss-Creations-Media-Kit-2020.pdf

Richardson, K. (2016a). Sex robot matters: Slavery, the prostituted, and the rights of machines. *IEEE Technology and Society Magazine, 35*(2), 46–53. https://doi.org/10.1109/MTS.2016.2554421

Richardson, K. (2016b). The asymmetrical 'relationship': Parallels between prostitution and the development of sex robots. *ACM SIGCAS Computers and Society, 45*(3), 290–293. https://doi.org/10.1145/2874239.2874281

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

Scheutz, M., & Arnold, T. (2017). Intimacy, bonding, and sex robots: Examining empirical results and exploring ethical ramifications. In J. Danaher & N. McArthur (Eds.), *Robot sex* (pp. 247–260). MIT Press.

Somerville, S. B. (1994). Scientific racism and the emergence of the homosexual body. *Journal of the History of Sexuality, 5*(2), 243–266.

Sparrow, R. (2017). Robots, rape, and representation. *International Journal of Social Robotics, 9*, 465–477. https://doi.org/10.1007/s12369-017-0413-z

Sparrow, R., & Karas, L. (2020). Teledildonics and rape by deception. *Law, Innovation and Technology, 12*(1), 175–204. https://doi.org/10.1080/17579961.2020.1727097

Westerman, D., Edwards, A. P., Edwards, C., Luo, Z., & Spence, P. R. (2020). I-It, I-Thou, I-Robot: The perceived humanness of AI in human-machine communication. *Communication Studies, 71*(3), 393–408. https://doi.org/10.1080/10510974.2020.1749683

Wilkerson, A. (2011). Disability, sex radicalism, and political agency. In K. Q. Hall (Ed.), *Feminist disability studies* (pp. 193–217). Indiana University Press.

Zatz, N. D. (1997). Sex work/sex act: Law, labor, and desire in constructions of prostitution. *Signs: Journal of Women in Culture and Society, 22*(2), 277–308. https://doi.org/10.1086/495157

Zhou, Y., & Fischer, M. H. (Ed.). (2019). *AI love you: Developments in human-robot intimate relationships*. Springer.

Zhou, Y., & Fischer, M. H. (2020). Intimate relationships with humanoid robots. In O. Bendel (Ed.), *Maschinenliebe: Liebespuppen und Sexroboter aus technischer, psychologischer und philosophischer Perspektive* (pp. 237–254). Springer.

I Get by With a Little Help From My Bots: Implications of Machine Agents in the Context of Social Support

Austin J. Beattie[1] and Andrew C. High[2]

1 Department of Communication Studies, University of Iowa, Iowa City, IA, USA
2 Department of Communication Arts and Sciences, The Pennsylvania State University, State College, PA, USA

Abstract

In this manuscript we discuss the increasing use of machine agents as potential sources of support for humans. Continued examination of the use of machine agents, particularly chatbots (or "bots") for support is crucial as more supportive interactions occur with these technologies. Building off extant research on supportive communication, this manuscript reviews research that has implications for bots as support providers. At the culmination of the literature review, several propositions regarding how factors of technological efficacy, problem severity, perceived stigma, and humanness affect the process of support are proposed. By reviewing relevant studies, we integrate research on human-machine and supportive communication to organize, extend, and provide a foundation for the growing body of work on machine agents for support.

Keywords: supportive communication, social support, human-machine communication, artificial intelligence, chatbots

At a nursing home in Michigan during the start of the COVID-19 pandemic, one resident called out "help me, I am in pain, I have to find a way to relieve it." This plea was not to staff or a family member, but to Amazon's Alexa. According to transcripts, the resident spent hours talking with Alexa through their quarantine and sought help over 40 times before passing away (Vigdor, 2020). With machine agents serving in more contexts than

CONTACT   Austin J. Beattie • Becker Communication Studies Building • University of Iowa • Iowa City, IA 52242-1498, USA •
austin-beattie@uiowa.edu

151

ever before, this story serves as an example of a phenomenon in human-machine communication (HMC) whereby humans engage digital interlocutors to seek and receive resources when they believe they need assistance.

Beyond people's informal use of relatively mainstream digital assistants like Alexa or Siri (e.g., Nedd, 2021), some machine agents have been designed specifically for human social and psychological needs. For instance, Joseph Wizenbaum's "Eliza" (the first chatbot circa 1966) and more contemporary examples like Stanford's "Woebot" (Fitzpatrick et al., 2017) specialize in Rogerian and Cognitive-Behavioral therapy, respectively. In response to these trends, researchers have begun to examine how people interact with machines in support contexts (e.g., Abendschein et al., 2021; Fitzpatrick et al., 2017; Kee et al., 2021; Rains et al., 2019a & 2019b). Support in the context of HMC may take many forms with machine-agents ranging from physical and embodied robots to internet-based chatbots. Throughout this manuscript, we recognize humans as support seekers and receivers who utilize machines as support providers. Additionally, just as there are effective and ineffective supportive interactions in both face-to-face and computer-mediated contexts when talking with other humans (High & Solomon, 2014; MacGeorge et al., 2011), we do not conceptualize machine agents to be exclusively successful support providers. Although machine agents may be more available than other humans for a potentially supportive interaction, users experience frustration when chatbots do not understand a user's commands or are unable to perform desired action (e.g., Abendschein et al., 2021). Due to their relative accessibility, this manuscript centers primarily on communicative support from internet chatbots or "bots" (e.g., Woebot) as opposed to embodied physical robots (e.g., Paro, Pepper).

This manuscript is an attempt to extend this growing body of research by organizing the existing scholarship and providing a foundation for research on supportive communication with bots. We review relevant research and provide propositions to guide future research at the culmination of our review. Although the extent to which machine agents as support providers generate substantive changes to the process of supportive communication is unclear, we follow prior research on supportive communication that demonstrates changes in supportive interactions based on differences in support providers (MacGeorge et al., 2011). Differences in supportive interactions between human or machine providers might be most common when people lack experience with HMC, and prior research indicates that scripts for interaction with bots are still evolving (e.g., Gambino et al., 2020). The benefits of support are contingent on effective interactions, which require the coordination of a support seeker, support provider, and numerous contextual factors that may vary through technology (MacGeorge et al., 2011).

This literature review first describes HMC in the context of support and then organizes HMC and related research that has implications for the processes of support seeking and provision. We center our review on empirical studies in HMC and related fields that examined variables predicting communication quality, such as how expressive people are (e.g., Mou & Xu, 2017) and how uncertain they feel before or after interacting with machines (e.g., A. Edwards et al., 2019; Spence et al., 2014). After we review this research, we describe its implications for future studies and present propositions regarding when machine agents might impair or improve supportive interactions to guide that research. In doing so, we provide a more nuanced understanding of HMC when communicating support and propose

whether and why machine agents might improve (or impair) supportive interactions relative to humans.

Supportive Communication

Scholars frequently describe social support as an interactive process. Supportive communication is verbal and nonverbal behavior that is enacted with the intention of helping others when they are perceived to require assistance (MacGeorge et al., 2011). Although scholars consider several distinct types of support, we focus our analysis on emotional support, which is the most commonly desired form of support (Burleson, 2008) and applicable across a range of stressors (MacGeorge et al., 2011). The validation and focus on affect that are inherent to emotional support are also useful to highlight some potentially influential differences between human and machine agent support providers.

High quality supportive communication bestows many benefits upon support seekers; however, such benefits are contingent on the successful coordination of the process of support. Both the seeking and provision of support vary in quality, and each phase of supportive communication is influenced by what happens previously (Barbee & Cunningham, 1995). Within supportive communication, typical studies are designed and limited to human support providers. Scholarship on HMC has also yet to seriously consider nuances within the process of supportive communication, despite the growing intersections of these bodies of research. Because the best understanding of supportive interactions with machines is perhaps achieved by taking seriously aspects of both technology and the process of supportive communication, this manuscript provides a framework to consider how support is sought and processed when engaging with a bot support provider.

Conceptualizing Supportive Communication With Machines

Due to the interactive nature of contemporary technology, the study of HMC, defined by Guzman (2018) as the "creation of meaning among humans and machines," has grown (p. 1). By incorporating both a social robotics (i.e., technology-based) and communicative (i.e., user-based) perspective, Rodríguez-Hidalgo (2020) argued that HMC presents "both perceived and enacted possibilities for social interaction in a two-way iterative communication process" (p. 62). Along the same lines, researchers interested in supportive communication differentiate support that is enacted from what is perceived, and in this manuscript, we often privilege support recipients' perceptions because they commonly maintain stronger associations with outcomes (MacGeorge et al., 2011). We conceptualize HMC in the context of support as the examination and practice of human-machine meaning-making and interaction from a perspective grounded in the intricacies of supportive communication to understand, predict, and explain how people communicate with machine agents when in need of aid. Like support in human-human interactions, we recognize that the goal of supportive interactions with machines is to receive effective assistance; however, when enacted, support can promote both positive and negative outcomes depending on how it is communicated and the source of the messages.

The technologies of human-machine communication in the context of support are limited only to those that users *perceive* to be socially interactive. In this way, we borrow from Gambino et al.'s (2020) conceptualization of *media agent* in general contexts as "any technological artifact that demonstrates sufficient social cues to indicate the potential to be a source of social interaction" (p. 73). In practice, this may include both sophisticated and basic programs and machinery ranging from chatbots developed for cognitive behavioral therapy, mood tracking, strategy games, and conversations with their users (see review by Abd-Alrazaq et al., 2020) to technologies like smart speakers that play songs, tell jokes, and serve as interaction partners (e.g., Gewirtz, 2016) in this conceptualization.

We align core foci from supportive communication to an HMC context, such as centering inquiry on the interactions between users and machines (support seekers and support providers, respectively), how people might seek help, and factors of quality support. The basic roles (i.e., seekers, providers) and processes (e.g., seeking, provision, etc.) of supportive communication remain unchanged in HMC, with the distinction of the provider role being occupied by a machine. Because providers are integral components of supportive interactions, switching the nature of the support provider leads to speculation about whether and how machine agents as support providers change the process of supportive communication. Because it is often the "first act" in the process of supportive communication, we consider how machine agents shape the process of seeking support before discussing how recipients process the messages they receive.

HMC and Seeking Support

Understanding how support is sought is important because the way people seek help influences the likelihood of support provision and its quality (Barbee & Cunningham, 1995; Cutrona, 1996). Seeking support is not a simple task, however, and tensions between goals of fully addressing a need for assistance while also managing face concerns (e.g., such as when a stressor may be perceived as embarrassing) complicate how people seek support (Goldsmith, 1995). Researchers commonly distinguish between whether seeking support occurs directly or indirectly, and Barbee and Cunningham noted that the costlier people perceive seeking support to be, the more indirect their strategies for seeking support will become. Direct seeking behaviors explicitly ask for help, signal a seeker's affect, and communicate an interest in addressing a problem, whereas indirect behaviors may minimize a problem or change the topic (Goldsmith, 1995; High & Scharp, 2015). Researchers generally argue that directly seeking support is most effective (Williams & Mickelson, 2008); however, it remains unclear if the presence of a machine agent would elicit more effective support seeking compared to a human.

Some research indicates that people are less interactive and expressive with bots and feel more uncertainty prior to interacting with them compared to humans (e.g., Mou & Xu, 2017; Spence et al., 2014), which suggests that people might seek support ineffectively from them. If seekers fail to engage in interactive and expressive behaviors with machines, the clarity and directness of their support seeking will likely suffer. In contrast, if bots can create an environment that fosters feelings of closeness or reduces costs related to seeking support, managing stigma, or accessing certain information (e.g., Fitzpatrick et al., 2017), people may seek support confidently and directly from them. In other words, by affording

users the ability to manage information with greater control while possibly feeling less judgment in their attempts to garner support compared to a human (Wright et al., 2010), bots might facilitate seeking support effectively. Based on these advantages and disadvantages, research provides mixed evidence regarding how bots affect the process of seeking support compared to humans.

HMC and Processing Support

After support is sought and provided, people evaluate the messages they receive, whether they are provided by machine agents or humans. Scholars assert that emotional support messages are often the most effective type of support regardless of the stressor, and the effectiveness of emotional support messages is determined by the extent to which they are verbally person-centered (VPC; Burleson, 2008; High & Dillard, 2012). VPC concerns how much a message illustrates "awareness of and adaptation to the subjective, affective, and relational aspects of communication contexts" and often predicts how effective (or ineffective) supportive messages are (High & Dillard, 2012). Rains et al. (2019a) found that chatbots providing high VPC messages received more positive evaluations than AI that provided low VPC messages. A question that remains concerns whether messages with a similar level of VPC produce equivalent outcomes when they are communicated by a human or a machine agent.

Understanding what makes support received during HMC effective is important not only from a practical standpoint, but also to advance how existing models of supportive communication explain, predict, or bring about further understanding of the effects of machine support providers. Myriad studies have reported a significant and positive relationship between VPC messages and positive support outcomes (High & Dillard, 2012); however, extant research is limited in its ability to explain or predict how varying degrees of person-centeredness will affect the process of support with a bot. Although Rains et al. (2019a) and Ho et al. (2018) demonstrated a bot's ability to provide some degree of person-centeredness, their findings do not unpack potential interactions between different support providers and supportive messages because they do not test a human vs. bot comparison. Perceptions of VPC messages on support outcomes are influenced by provider characteristics (e.g., High & Solomon, 2014), and exploring how outcomes vary between human and machine providers will advance studies that consider elements of technology that elicit the best support outcomes (High & Solomon, 2011).

Some research suggests that high VPC messages from a bot will be perceived unfavorably compared to the same messages from a human. When describing supportive interactions, Applegate (1980) wrote that "abstract, dispositionally oriented constructs for perceiving others results in more stable and individually adapted impressions for formulating listener-adapted communication messages" (pp. 61–62). This conception is situated in assumptions of human emotion, cognition, and expression. The inability to feel, think, or speak quite like another human may make differences between a human and robot more apparent when communicating emotional or personal messages. Researchers have found people are sometimes less verbally responsive and emotionally expressive when talking to machines (Kanda et al., 2008; Mou & Xu, 2017; Prahl & Van Swol, 2021). Providing the highest levels of person-centered messages requires cognitive complexity, relational history,

and the ability to tailor individualized messages (Burleson, 1982); therefore, perceptions of messages from bots, even if they contain objectively effective content, may still fall behind those of messages created by humans.

Like their relative ineffectiveness from human providers, low VPC messages from a machine agent might yield particularly negative or unfavorable outcomes. Low VPC messages are thought to be worse when communicated online than in person (High & Solomon, 2014), suggesting that technological contexts may yield worsened impressions of low VPC messages. Extending this line of thinking, because people are less satisfied with low compared to high VPC messages (and may be even less so when communicated online) to begin with, they may perceive them to be even worse when coupled with an unnatural support provider (e.g., Lee, 2004).

Other research suggests that machine agents communicating VPC messages will improve support outcomes and be rated more favorably than those from a human. Evaluations of messages are shaped by both message and source characteristics, and bots may facilitate a "weak tie" relationship that might benefit supportive interactions. Weak ties do not require obligation to reciprocate support and often correspond with less judgment compared to interactions with closer contacts (Wright et al., 2010). These advantages to receiving support from comparatively weak ties might benefit supportive interactions with machine agents.

Message sophistication also influences how people evaluate machine agents. For example, using a message design logic framework (O'Keefe, 1988), A. Edwards and colleagues (2020) documented that a humanoid robot employing rhetorical logics (i.e., more sophisticated messages that are flexible and address multiple goals) was met with higher ratings of credibility, attractiveness, and competence from participants compared to when it employed less-sophisticated messages. One participant remarked the robot employing sophisticated messages was "very understanding, more than I would be" and that "[the robot] may genuinely care about the group members" (Edwards et al., 2020, p. 953), indicating people may prefer more-sophisticated messages from machines and perhaps suggesting machines may be able to provide those messages more effectively than other humans.

If findings from Edwards et al. (2020) extend to supportive exchanges, and people are more comfortable talking to machines that provide sophisticated messages than they are with other humans, the least effective combination of provider and level of VPC might be low person-centered messages from human providers. Central to this scenario is the assumption that low VPC messages, characterized by ignoring or deflecting the concerns of others (Burleson, 2008), might seem more hurtful when coming from a human (who might "know better") than a machine. In other words, such violations might be attributed to technological errors or limitations, rather than potentially face-threatening acts that might come from a human.

Rains et al. (2019a) advanced understanding of support in HMC by demonstrating that perceptions of VPC vary within messages communicated by bots, but whether and how these perceptions are altered between bots (or other machine agents) and human support providers has yet to be fully understood. Without a human control-group, the idea of equivalency or a direct comparison between human and machine support cannot be fully examined. According to Bodie and Burleson (2008), "enhancing the success of helpers who provide support requires a comprehensive explanation of why support messages are

effective in some circumstances but less effective in others" (p. 355). Identifying whether the same messages are evaluated differently in contexts of HMC helps to understand the influential elements of supportive conversations, thereby potentially leading to more satisfying support outcomes.

Does HMC Impair or Improve Processes of Supportive Communication?

To this point, we provided a general description of supportive communication and its associated processes, and we conceptualized the role of bots in the process of supportive communication. We also briefly described how bots are implicated in the processes of support seeking and message processing. Generally, this review of the literature presents mixed results. Some research suggests that machine agents impair supportive interactions (e.g., Mou & Xu, 2017). In contrast, other research suggests that bots have the capacity to improve supportive communication (e.g., Ta et al., 2020). Based on these mixed results, we look more closely at studies that suggest bots can impair or improve supportive interactions, focus on the implications of those studies for seeking and processing supportive messages, and consider factors that might determine whether it is more likely that machine agents will impair or improve processes of support. Doing so highlights the implications of our literature review, provides testable propositions for future research, and establishes a foundation for research on HMC in the context of supportive communication. One line of research, which we refer to as the *impairment* perspective, generally suggests people will respond unfavorably to machines compared to humans in supportive interactions. Broadly speaking, the impairment perspective is represented by general attitudes, opinions, and evidence that suggest supportive interactions with machines will be inferior to those between two humans.

The *impairment* perspective is based, in part, on the premise that people's expectations and scripts for interaction are meant for other humans, not machines. For instance, scholars have argued that people are driven by evolutionary (e.g., Lee, 2004) or ontologically-based classifications (e.g., Bolter, 1984) to interact with other humans compared to machine agents. This perspective is further buttressed by applied studies where machines harm processes that are critical to the communication of support. For example, Kanda et al. (2008) found that people were less nonverbally responsive to physical robots than other humans, arguing people felt aversion toward the robots. In another study on physical robots, people also reported lower satisfaction and intent for future interaction, especially when they felt stressed prior to talking with the robot (Ling & Björling, 2020). Mou and Xu (2017) documented that people were less open, agreeable, extroverted, and self-disclosive with chatbots than humans, further suggesting that machine agents may not make attractive partners for support.

The notion that machines might impair supportive communication is further backed by issues surrounding a machine's human likeness or task ability. Keijsers and Bartneck (2018) asked participants to interact with a digital representation of a physical robot and found that reducing the nonverbal cues the robot communicated was associated with higher levels of participant aggression. Humans also rate other humans higher in perceived expertise, usability, and similarity in decision-making tasks compared to machines (Prahl & Van Swol, 2021). Considering these findings, Ho et al. (2018) presented a "perceived

understanding" framework, which argued that because humans are perceived to have more empathy than machines, emotional, relational, and psychological effects of support will be greater when disclosing to a person than to a bot. Because clear communication is critical to effective support seeking and provision (e.g., Williams & Mickelson, 2008), if bots cause support seekers to feel they are interacting with a less useful partner, support processes may suffer.

Expectancies for interaction and whether people feel they are understood are consequential to processes of social support (MacGeorge et al., 2011). People are not likely to seek or positively evaluate support from an entity with whom they feel uncertain, do not like, or feel distant, effects that have been observed in various studies on human-machine interaction (e.g., A. Edwards et al., 2019; C. Edwards et al., 2016; Spence et al., 2014). People are also unlikely to seek help from machine agents if they are less agreeable with bots (e.g., Mou & Xu, 2017). Often based on expectations, scripts, number of cues, or amounts of presence, multiple studies suggest that machines harm communication. Accordingly, we extend that research and offer a general impairment hypothesis in the context of supportive communication:

> **Impairment Perspective:** Machine agent support providers have a negative effect on support processes and outcomes compared to human support providers.

The *improvement* perspective is rooted in research suggesting that machine agents present an overall positive influence on social support. The improvement perspective borrows from literature citing the benefits of asynchronous and reduced-cue environments for supportive interactions and hyperpersonal effects (e.g., Walther, 1996; Walther & Boyd, 2002). This perspective also notes that machines possess the capacity to convey increasingly complex emotional messages that recipients perceive to be tailored to them and their needs (e.g., Ho et al., 2018; Rains et al., 2019a). It also argues that new scripts and boundaries between people and machines are emerging and evolving (e.g., A. Edwards et al., 2019; Gambino et al., 2020; Guzman, 2018). Central to this perspective is the idea that machines may provide advantages for supportive communication over traditional human-human supportive interactions.

Issues of access, adequacy (e.g., Walther & Boyd, 2002), anonymity (e.g., Rheingold, 1993), and stigma (e.g., Williams et al., 2016; Williams & Mickelson, 2008) impact whether and how people seek support. Communicating via technology has been observed to manage these considerations in beneficial ways, perhaps especially stigma, and communicating *with* technology might involve similar benefits. Albrecht and Adelman (1987) stated eloquently that CMC facilitates "low-risk discussions about high-risk topics" (p. 133). Because most bots designed for support lack either sophisticated representations or material embodiment, their expression is discernable enough to perceive friendliness and warmth but suppressed to the point that participants feel free to express themselves. In other words, machine agents might promote disinhibition and anonymity effects that foster fuller disclosure and positive support provision (e.g., Fitzpatrick et al., 2017; Ho et al., 2018).

Beyond facilitating seeking support, machine agents may also be able to provide support that is enriching and effective. For example, Ta et al. (2020) found that chatbots serve as

an important source of companionship in addition to emotional, informational, and esteem support for their users. Leite et al. (2012) found that children perceive informational support from robots in their classroom similar to support from their peers, an effect they argued was due to the bots' ability to convey messages that were empathetic and encouraging. Both task-based and social interactions with machine agents foster trust, and those positive attitudes improve as experience with them increases (Banks et al., 2021). Ho et al. (2018) had participants disclose informational or emotional content to chatbots or humans and found no significant differences in perceptions of relational warmth or closeness between conditions. Beattie et al. (2020) found that impressions of interpersonal attractiveness and credibility were higher for humans and chatbots that used emoji than those that employed verbal-only messages. Furthermore, in a two-part study featuring chat transcripts and live chatbot interactions, Liu and Sundar (2018) found that participants favored when bots provided sympathy and empathy over unemotional provision of advice.

People may also be developing scripts specifically for interaction with machine agents. For instance, Gambino and colleagues (2020) argued that "given a deeper and broader realm of experience [with bots], humans may implement scripts they have developed for interactions specific to media entities." (p. 72). A. Edwards et al. (2019) found that participants report less uncertainty following an interaction with a humanoid robot compared to before the interaction, which they reasoned was due to participants adjusting their expectations for interpersonal cues (e.g., nonverbal confirmation) based on a consideration of the technological limitations of the robot. These findings support the notion of a machine-specific interaction script. C. Edwards et al. (2016) suggested that with adequate time, people form interaction scripts specifically for HMC that mirror hyperpersonal effects. When interacting or utilizing scripts for interactions with machine agents, people might focus less energy worrying about their partner's feelings or paying attention to their body language and instead invest more energy on carefully crafting and preparing their own messages (e.g., cognitive reallocation; Walther, 1996). Bots might also represent a source of weak tie support, which is defined as relationships that people maintain but are not intimately close. Compared to strong ties, like many typical human support providers, weak ties benefit supportive interactions because they involve reduced risk, a greater variety of information, less judgment, and less role obligation (Wright et al., 2010).

In sum, the improvement perspective is represented by arguments and findings that suggest the potential for satisfying supportive outcomes with machines. Scholars argue that machine agents create support that helps people reappraise their problems (e.g., Rains et al., 2019a), conveys emotionally rich messages (e.g., Ho et al., 2018), and affords users greater anonymity and stigma management (e.g., Fitzpatrick et al., 2017). Machine agents might foster the creation of new schema that can facilitate supportive processes. The culmination of these studies is thinking related to the improvement perspective, which asserts that machine agents can benefit supportive communication:

> **Improvement Perspective**: Machine agent support providers have a positive effect on support processes and outcomes compared to human support providers.

When Will the Impairment or Improvement Perspective Be Experienced?

The impairment and improvement perspectives offer competing predictions for how machines will affect processes of supportive communication. Due to the varied nature of interpersonal communication and support, both perspectives may exert their effects at different times. We review some relevant variables that might determine when machine agents impair or improve supportive interactions below and present several propositions specifying when each perspective is likely to be relevant. Although not exhaustive, we focus on whether a user is comfortable with technology, how severe or stigmatizing they perceive a stressor to be, and the human social cues present in a machine in the following pages. Further research will help forecast additional contingencies that determine when either perspective will result.

Comfort, competence, or efficacy with technology is likely a predictor of whether machines will impair or improve supportive interactions. For instance, many senior citizens own smartphones but report difficulty using them (Jefferson, 2019), and their lack of efficacy discourages them and impairs their use. Individuals who are less comfortable using technology and think they are less effective doing so generally report less use of technology (Caplan, 2003). In contrast, people who report high levels of technological efficacy generally use more channels of communication and use them more frequently than people who are less efficacious (LaRose et al., 2003). Some scholars even observe that the use of online tools is a form of self-efficacy or skill building that can restore control over serious issues (Rottmann et al., 2010). In much the same way, we expect people with higher levels of technological efficacy to take advantage of the benefits of support from machine agents. Based on this logic, we propose:

> **Proposition 1**: Greater levels of technological efficacy correspond with HMC improving more than impairing supportive interactions and outcomes.

Aspects of a person's stressor such as its severity and perceived stigma are other important factors that likely determine whether bots impair or improve supportive interactions. Stressors vary in terms of magnitude, permanence, and how stigmatizing they are, and HMC might be an impairment for major issues. People's likelihood of seeking support increases as the severity of a stressor increases (Oh & LaRose, 2016), but they might prefer the expertise of another human for serious problems. High VPC messages are also most effective in severe situations (Bodie, 2013), yet as previously discussed, machines might have difficulty producing those messages. Although technology is improving, people generally prefer affect-oriented or nurturant support for severe issues (Rains et al., 2015), which might be more natural from humans than bots. This research suggests that people avoid bots when coping with severe stressors, which leads to the following prediction:

> **Proposition 2**: Greater levels of problem severity correspond with HMC impairing more than improving supportive interactions and outcomes.

The perceived stigma associated with a stressor is another factor that might determine whether HMC impairs or improves supportive exchanges. People who feel stigmatized are often reluctant to seek support from others and do so in ineffective or indirect ways because

they fear rejection from potential support providers (Williams & Mickelson, 2008). Of course, this stigma comes from other humans, and the lack of judgment within online venues makes them popular for discussing stigmatizing issues (Bargh & McKenna, 2004). Stigma promotes the impression that support is unavailable (Mickelson & Williams, 2008), but such concerns are expected to be assuaged with bots that are constantly available. The ease of communicating about a stigmatized stressor might be enhanced when interacting with a nonjudgmental, anonymous machine agent. Along these lines, people coping with a stigma disclose more in anonymous contexts when they do not feel like they are being judged than when anonymity is low (Rains, 2014). DeAndrea (2015) documented that stigma compels people to use online compared to in-person support groups, and we extend this thinking to suggest that the same benefits improve supportive interactions with bots. Hence:

> **Proposition 3**: Greater levels of perceived stigma correspond with HMC improving more than impairing supportive interactions and outcomes.

The number of human-like cues present in a bot might also improve supportive interactions. As previously stated, the fewer anthropomorphic cues a robot communicates, the more aggressive participants are when interacting with it (Keijsers & Bartneck, 2018). People also feel less uncertainty and higher social presence after initial interactions with a humanoid robot compared to a human (A. Edwards et al., 2019). These results indicate that adding human-like cues or enhancing the social presence in an interaction with machines can reduce negative expectations or effects of interacting with them. There is also a positive association between the anthropomorphism of a chatbot and its social presence and conversational skill (Schuetzler et al., 2020). Given research that suggests humans value and respond effectively to anthropomorphic machines, people might desire a bot that facilitates enough human cues to be understandable, but perhaps one that lacks the judgment and faces threats inherent to supportive interactions with humans. Although there appears to be a point when too many human cues produce an uncanny valley effect that reduces positive social outcomes, the type of conversational bots that are the focus of this analysis are unlikely to reach that level of humanness. From this logic, we propose:

> **Proposition 4**: Greater levels of human social cues in HMC improve more than impair supportive interactions and outcomes.

Future Considerations and Conclusion

As research continues and technology becomes more sophisticated, factors that could impair or improve HMC in the context of support are likely to exhibit concomitant evolution. Although we grouped the studies we reviewed along the lines of impairment or improvement to synthesize relevant research on support and HMC, more systematic examination of extant research and further empirical study are needed to produce more nuanced and sophisticated understandings of how machine agents influence the support process. The characteristics and qualities of contexts in which machines may impair or improve supportive outcomes, as well as how factors such as technological efficacy, the severity of the stressor, how stigmatizing a stressor is perceived to be, or a bot's degree of humanness

influence the process of support represent several clear starting points for further inquiry. Along these lines, a machine's ability to process natural language (e.g., vocal tone, pitch, and behavior patterns) further suggests that bots may be more perceptive of human emotional states in the future. Such advancements may obscure distinctions between impairment and improvement or lead to new perspectives altogether. Although some current iterations of natural language processing are fairly crude, especially in regard to variations in slang or dialect (Zou & Schiebinger, 2018) and might exhibit gender bias (Sun et al., 2019), if successful, these technologies will likely influence the scripts people hold for HMC. These improvements might even allow machine agents to substitute, rather than complement humans, thereby relegating humans to secondary support providers.

A potentially influential difference between supportive interactions with another human versus a machine is that the roles of support provider and seeker/receiver shift and are reciprocated over time when conversing with another human. In HMC, the provision of support will almost always flow from the bot to the human, negating the opportunity to reciprocate support. Reciprocating support provides many benefits (MacGeorge et al., 2011); for instance, people report the most comfort receiving support when they rationalize that they provided substantial amounts of support to others previously (Kuijer et al., 2001). Less is known about the relational dynamics or outcomes when someone is a constant source of unreciprocated support for a partner. Although human caregivers become burned out (Harvey-Knowles & Faw, 2017), bots do not fatigue in the same way humans do. Future research can examine if receiving support from machines without the opportunity to reciprocate diminishes the efficacy of support changes how people seek support from them or creates new ways of reciprocating support that are unique to support in HMC contexts. For example, Meng and Dai (2021) examined reciprocity of self-disclosure in machine support and found that self-disclosure from a bot (i.e., sharing a "worry" the bot has, such as fear of a technical malfunction) combined with supportive messages was positively associated with participant's worry reduction for their own stressors, suggesting reciprocity still plays a vital role in human-machine support contexts. Perhaps the maintenance and updates that are required by technology can be viewed as a means of reciprocating support.

These directions for research are far from exhaustive. Perhaps the most consistent quality of human behavior and technology is the continual change and accompanying questions they bring for researchers and end-users alike to disentangle and determine when interactions are improved, impaired, or changed to something entirely new. Fortunati and Edwards (2020) asked "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?" (p. 9). Further study can address these questions to inform HMC scholarship, designers of interactive systems, and, most importantly, the people whose supportive interactions stand to be improved (or impaired) through HMC. We return to the Michigan resident who talked to Alexa in their final hours. With most of the world in quarantine, Alexa was "ready" to help. Although we will never know if the patient was satisfied or comforted by the help they received, the story suggests a shift in how people will seek help in the future. We hope improving such interactions and others like it will be accomplished through further and more rigorous examination of supportive communication in HMC.

Acknowledgment

Author Biographies

Austin J. Beattie (MA, Western Michigan University) is a PhD Candidate at the University of Iowa and a member of the Communication and Social Robotics Labs (www.combot-labs.org). His research focusses on social support and community engagement in computer-mediated and human-machine communication. In particular, Austin's dissertation work examines how variations of message source and sophistication influence the ways in which social support is engaged, facilitated, and evaluated in HMC contexts.

  https://orcid.org/0000-0003-2667-4321

Andrew C. High (PhD, The Pennsylvania State University) is an associate professor in the Department of Communication Studies at The Pennsylvania State University. His research interests focus on supportive communication and computer-mediated communication. In particular, his research investigates personal, contextual, and relational qualities that influence the processes of seeking, providing, and receiving support, in addition to whether and how different channels or features of computer-mediated communication influence the process of support.

  https://orcid.org/0000-0002-9397-6875

References

Abd-Alrazaq, A. A., Rababeh, A., Alajlani, M., Bewick, B. M., & Househ, M. (2020). Effectiveness and safety of using chatbots to improve mental health: Systematic review and meta-analysis. *Journal of Medical Internet Research*, *22*(7), e16021. https://doi.org/10.2196/16021

Abendschein, B., Edwards, C., & Edwards, A. (2021). The influence of agent and message type on perceptions of social support in human-machine communication. *Communication Research Reports, 38*(5), 304–314. https://doi.org/10.1080/08824096.2021.1966405

Albrecht, T. L., & Adelman, M. B. (Eds.). (1987). *Communicating social support*. Sage Publications, Inc.

Applegate, J. L. (1980). *Person-centered and position-centered teacher communication*. In N. K. Denizen (Ed.), *Studies in symbolic interaction* (pp. 59–96) Vol. 3. JAI Press.

Banks, J., Koban, K., & Chauveau, P. D. V. (2021). Forms and frames: Mind, morality, and trust in robots across prototypical interactions. *Human-Machine Communication*, *2*, 81–103. https://doi.org/10.30658/hmc.2.4

Barbee, A. P., & Cunningham, M. R. (1995). An experimental approach to social support communications: Interactive coping in close relationships. *Annals of the International Communication Association*, *18*, 381–413. https://doi.org/10.1080/23808985.1995.11678921

Bargh, J. A., & McKenna, K. Y. A. (2004). The internet and social life. *Annual Review of Psychology, 55,* 573–590. https://doi.org/10.1146/annurev.psych.55.090902.141922

Beattie, A., Edwards, A. P., & Edwards, C. (2020). A bot and a smile: Interpersonal impressions of chatbots and humans using emoji in computer-mediated communication. *Communication Studies*, 1–19. https://doi.org/10.1080/10510974.2020.1725082

Bodie, G. D. (2013). The role of thinking in the comforting process: An empirical test of a dual-process framework. *Communication Research*, *40*, 533–558. https://doi.org/10.1177/0093650211427030

Bodie, G. D., & Burleson, B. R. (2008). Explaining variations in the effects of supportive messages a dual-process framework. *Annals of the International Communication Association*, *32*, 355–398. https://doi.org/10.1080/23808985.2008.11679082

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

Burleson, B. R. (2008). What counts as effective emotional support? Explorations of individual and situational differences. *Studies in Applied Interpersonal Communication*, 207–227. Sage Publications, Inc. https://doi.org/10.4135/9781412990301.d14

Caplan, S. E. (2003). A social skill account of problematic internet use. *Journal of Communication*, *55*, 721–736. https://doi.org/10.1111/j.1460-2466.2005.tb03019.x

Cutrona, C. E. (1996). Social support in couples: Marriage as a resource in times of stress. SAGE.

DeAndrea, D. C. (2015). Testing the proclaimed affordances of online support groups in a nationally representative sample of adults seeking mental health assistance. *Journal of Health Communication, 20*(2), 147–156. https://doi.org/10.1080/10810730.2014.914606

Edwards, A., Edwards, C., & Gambino, A. (2020). The social pragmatics of communication with social robots: Effects of robot message design logic in a regulative context. *International Journal of Social Robotics*, *12*, 945–957. 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

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

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health*, *4*, e19. https://doi.org/10.2196/mental.7785

Fortunati, L., & Edwards, A. P. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, *1*, 1. https://doi.org/10.30658/hmc.1.1

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*, 5. https://doi.org/10.30658/hmc.1.5

Gewirtz, D. (2016). Seven secret Alexa tricks that will help you wake up, cheer up, and settle a toss up. *ZDNET*. https://web.archive.org/web/20210518110700/https://www.zdnet.com/article/seven-secret-alexa-tricks-that-will-help-you-wake-up-cheer-up-and-settle-a-toss-up/

Goldsmith, D. J. (1995). The communicative micro-dynamics of support. *Annals of the International Communication Association*, *18*, 414–433. https://doi.org/10.1080/23808985.1995.11678922

Guzman, A. (2018). What is human-machine communication anyways. In A. Guzman (Ed.), *Human-machine communication: Rethinking communication technology and ourselves* (pp. 1–28). Peter Lang.

Harvey-Knowles, J., & Faw, M. H. (2017). Caregiver social support quality when interacting with cancer survivors: Advancing the dual-process model of supportive communication. *Supportive Care in Cancer*, 1–8. https://doi.org/10.1007/s00520-017-3951-x

High, A. C., & Dillard, J. P. (2012). A review and meta-analysis of person-centered messages and social support outcomes. *Communication Studies*, *63*, 99–118. https://doi.org/10.1080/10510974.2011.598208

High, A. C., & Scharp, K. M. (2015). Examining family communication patterns and seeking social support direct and indirect effects through ability and motivation. *Human Communication Research*, *41*, 459–479. https://doi.org/10.1111/hcre.12061

High, A. C., & Solomon, D. H. (2011). Locating computer-mediated social support within online communication environments. In K. B. Wright & L. M. Webb (Eds.), *Computer-mediated communication in personal relationships* (pp. 119–155). Peter Lang Publishing.

High, A. C., & Solomon, D. H. (2014). Communication channel, sex, and the immediate and longitudinal outcomes of verbal person-centered support. *Communication Monographs*, *81*, 439-468. https://doi.org/10.1080/03637751.2014.933245

Ho, A., Hancock, J., & Miner, A. (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

Jefferson, R. (2019). More seniors are embracing technology, but can they use it? UCSD researchers suggest asking them. *Forbes*. https://www.forbes.com/sites/robinseatonjefferson/2019/06/28/more-seniors-are-embracing-technology-but-can-they-use-it-ucsd-researchers-suggest-asking-them/

Kanda, T., Miyashita, T., Osada, T., Haikawa, Y., & Ishiguro, H. (2008). Analysis of humanoid appearances in human–robot interaction. *IEEE Transactions on Robotics, 24*, 725–735. https://doi.org/10.1109/IROS.2005.1544960

Kee, K. F., Calyam, P. P., & Regunath, H. (2021). The role of vidura chatbot in the diffusion of KnowCOVID-19 gateway. *Human-Machine Communication*, *3*(1), 4. https://doi.org/10.30658/hmc.3.4

Keijsers, M., & Bartneck, C. (2018). Mindless robots get bullied. *In Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction* (pp. 205–214). https://doi.org/10.1145/3171221.3171266

Kuijer, R. G., Buunk, B. P., & Ybema, J. F. (2001). Justice of give-and-take in the intimate relationship: When one partner of a couple is diagnosed with cancer. *Personal Relationships*, *8*(1), 75–92. https://doi.org/10.1111/j.1475-6811.2001.tb00029.x

LaRose, R., Lin, C. A., & Eastin, M. S. (2003). Unregulated internet usage: Addiction, habit, or deficient self-regulation? *Media Psychology*, *5*, 225–253. https://doi.org/10.1207/S1532785XMEP0503_01

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

Leite, I., Castellano, G., Pereira, A., Martinho, C., & Paiva, A. (2012). Long-term in- teractions with empathic robots: Evaluating perceived support in children. In S. Ge, O. Khatib, J.-J. Cabibihan, R. Simmons, & M.-A. Williams (Eds.), *Social robotics* (Vol. 7621, pp. 298–307). Springer. https://doi.org/10.1007/978-3-642-34103-8_30

Ling, H. Y., & Björling, E. A. (2020). Sharing stress with a robot: What would a robot say? *Human-Machine Communication*, *1*, 133-159. https://doi.org/10.30658/hmc.1.8

Liu, B., & Sundar, S. S. (2018). Should machines express sympathy and empathy? Experiments with a health advice chatbot. *Cyberpsychology, Behavior, and Social Networking*, *21*, 625–636. https://doi.org/10.1089/cyber.2018.0110

MacGeorge, E. L., Feng, B., & Burleson, B. R. (2011). Supportive communication. In M. L. Knapp & J. A. Daly (Eds.), *The SAGE handbook of interpersonal communication* (4th ed., pp. 317–354). Sage.

Meng, J., & Dai, Y. N. (2021). Emotional support from AI chatbots: Should a supportive partner self-disclose or not? *Journal of Computer-Mediated Communication*, *4*, 207–222. https://doi.org/10.1093/jcmc/zmab005

Mickelson, K. D., & Williams, S. L. (2008). Perceived stigma of poverty and depression: Examination of interpersonal and intrapersonal mediators. *Journal of Social and Clinical Psychology*, 27, 903–930. https://doi.org/10.1521/jscp.2008.27.9.903

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

Nedd, A. (2021). Apple discontinued the one thing that got me through the pandemic. *Mashable*. https://web.archive.org/web/20220105001210/https://mashable.com/article/apple-homepod

Oh, H. J., & LaRose, R. (2016). Impression management concerns and support-seeking behavior on social network sites. *Computers in Human Behavior*, *57*, 38–47.

O'Keefe, B. (1988). The logic of message design: Individual differences in reasoning about communication. *Communications Monographs*, *55*, 80–103. https://doi.org/10.1080/03637758809376159

Prahl, A., & Van Swol, L. (2021). Out with the humans, in with the machines? Investigating the behavioral and psychological effects of replacing human advisors with a machine. *Human-Machine Communication, 2*, 11. https://doi.org/10.30658/hmc.2.11

Rains, S. A. (2014). The implications of stigma and anonymity for self-disclosure in health blogs. *Health Communication*, *29*, 23–31. https://doi.org/10.1080/10410236.2012.714861

Rains, S. A., Pavlich, C. A., Lutovsky, B., Tsetsi, E., & Ashtaputre, A. (2019b). Support seeker expectations, support message quality, and supportive interaction processes and outcomes: The case of the comforting computer program revisited. *Journal of Social and Personal Relationships*, *37*, 647–666. https://doi.org/10.1177/0265407519876359

Rains, S. A., Pavlich, C. A., Tsetsi, E., & Akers, C. (2019a). Evaluating the scope and theoretical mechanisms of person-centered social support: The case of the comforting computer program. *Communication Research*, *47*, 512–540. https://doi.org/10.1177/0093650219852554

Rains, S. A., Peterson, E. B., & Wright, K. B. (2015). Communicating social support in computer-mediated contexts: A meta-analytic review of content analyses examining support messages shared online among individuals coping with illness. *Communication Monographs*, *82*, 403–430. https://doi.org/10.1080/03637751.2015.1019530

Rheingold, H. (1993). *The virtual community: Finding connection in a computerized world.* Addison-Wesley Longman Publishing Co., Inc.

Rodríguez-Hidalgo, C. (2020). Me and my robot smiled at one another: The process of socially enacted communicative affordance in human-machine communication. *Human-Machine Communication, 1*, 55–69. https://doi.org/10.30658/hmc.1.4

Rottmann, N., Dalton, S. O., Christensen, J., Frederiksen, K., & Johansen, C. (2010). Self-efficacy, adjustment style and well-being in breast cancer patients: A longitudinal study. *Quality of Life Research*, *19*, 827–836. https://doi.org/10.1007/s11136-010-9653-1

Schuetzler, R. M., Grimes, G. M., & Scott Giboney, J. (2020). The impact of chatbot conversational skill on engagement and perceived humanness. *Journal of Management Information Systems, 37*, 875–900. https://doi.org/10.1080/07421222.2020.1790204

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

Sun, T., Gaut, A., Tang, S., Huang, Y., ElSherief, M., Zhao, J., Mirza, D., Belding, E., Chang, K.-W., & Wang, W. Y. (2019). Mitigating gender bias in natural language processing: Literature review. *arXiv preprint arXiv:1906.08976*.

Ta, V., Griffith, C., Boatfield, C., Wang, X., Civitello, M., Bader, H., De Cero, E., & Loggarakis, A. (2020). User experiences of social support from companion chatbots in everyday contexts: Thematic analysis. *Journal of Medical Internet Research*, *22*. https://www.jmir.org/2020/3/e16235/

Vigdor, N. (2020). "How do I get help?" Dying coronavirus patient asked Alexa. *The New York Times*. http://web.archive.org/web/20200412083434/https://www.nytimes.com/2020/04/09/us/Coronavirus-Alexa-ask-for-help.html

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

Walther, J. B., & Boyd, S. (2002). Attraction to computer-mediated social support. *Communication Technology and Society: Audience Adoption and Uses*, *153188*, 50–88.

Williams, S. L., Laduke, S. L., Klik, K. A., & Hutsell, D. W. (2016). A paradox of support seeking and support response among gays and lesbians. *Personal Relationships*, *23*(2), 296–310. https://doi.org/10.1111/pere.12127

Williams, S. L., & Mickelson, K. D. (2008). A paradox of support seeking and rejection among the stigmatized. *Personal Relationships*, *15*, 493–509. https://doi.org/10.1111/j.1475-6811.2008.00212.x

Wright, K. B., Rains, S., & Banas, J. (2010). Weak-tie support network preference and perceived life stress among participants in health-related, computer-mediated support groups. *Journal of Computer-Mediated Communication*, *15*, 606–624. https://doi.org/10.1111/j.1083-6101.2009.01505.x

Zou, J., & Schiebinger L. (2018). AI can be sexist and racist—It's time to make it fair. *Nature*, 559, 324–326. https://doi.org/10.1038/d41586-018-05707-8

Embracing AI-Based Education: Perceived Social Presence of Human Teachers and Expectations About Machine Teachers in Online Education

Jihyun Kim[1] , Kelly Merrill Jr.[2] , Kun Xu[3] , and Deanna D. Sellnow[1]

1 Nicholson School of Communication and Media, University of Central Florida, Orlando, FL, USA
2 School of Communication, The Ohio State University, Columbus, OH, USA
3 Department of Media Production, Management, and Technology, University of Florida, Gainesville, FL, USA

Abstract

Technological advancements in education have turned the idea of machines as teachers into a reality. To better understand this phenomenon, the present study explores how college students develop expectations (or anticipations) about a machine teacher, particularly an AI teaching assistant. Specifically, the study examines whether students' previous experiences with online courses taught by a human teacher would influence their expectations about AI teaching assistants in future online courses. An online survey was conducted to collect data from college students in the United States. Findings indicate that positively experienced social presence of a human teacher helps develop positive expectations about an AI teaching assistant. The study provides meaningful implications and contributions to our understanding of a machine agent in education.

Keywords: artificial intelligence; AI instructor; machine teacher; online education; social presence

While the past decades benefitted from the use of technology as a tool that allows for the creation of an online learning environment and enhances learning outcomes in a traditional classroom (e.g., Al Ghamdi et al., 2016; Kaufmann et al., 2016; Limperos et al., 2015),

169

educators have recently started to use technology as an agent in the classroom. For example, Jill Watson, the world's first artificial intelligence (AI) teaching assistant, which was built based on IBM's Watson platform, was introduced to help students in an online learning environment. This AI teaching assistant helps students by answering questions in a course management website. The advent of Jill Watson signals that this new era of education is starting to incorporate machine teachers in education. The current status of technology-based education is already in the process of transforming to a new level (J. Kim, 2021).

Acknowledging the importance of technology in education, research has explored the idea of machine teachers in higher education. For example, A. Edwards and Edwards (2017) highlighted the importance of machines, such as robots, in effective learning. Further, J. Kim et al. (2020, 2021) examined college students' perceptions about an AI teaching assistant and AI instructor in online education. These studies provide preliminary but foundational understanding about the new era of education. However, there exists much more to explore in this realm.

Given the current trend of technology use in education, it is possible that universities may implement AI-based education at various levels in the coming future. One of the key differences between traditional online education and AI-based online education is the nature of the teacher, human versus machine. Then, the question is whether students' previous experiences with a human teacher in an online learning environment would translate into students' expectations (or anticipations) about a machine teacher, which could eventually affect their perceptions and adoption of an AI-based online education.

To address the above-mentioned question, we focus on one particular aspect, social presence. Social presence was selected because of its importance in creating positive online learning experiences (e.g., J. Kim et al., 2016; Song, Kim, & Park, 2019). Students can feel socially connected to their teacher (J. Kim et al., 2016; Song, Kim, & Park, 2019), and this highlights that students develop social perceptions of their teacher based on available cues regardless of whether they meet with the instructor in a physical classroom or not. In this regard, the study examines whether previously experienced social presence of a human teacher helps students develop certain expectations about a machine teacher, who students would not meet in a face-to-face context but would serve in a similar role as a human teacher. Specifically, the study focuses on an AI teaching assistant, which typically assists a primary instructor, such as answering questions from students about the course and the assignments on the course management website. The following section begins with a review of literature, followed by a method, results, and a discussion of the findings, implications, and contributions of the study.

Social Presence in Online Education

Given that social presence is an important experience in a mediated environment (Biocca et al., 2003), the notion has been understood and described from various perspectives and approaches (e.g., Biocca et al., 2003; Lee, 2004; Lombard & Ditton, 1997). Of them, Lee (2004) conceptualizes *social presence* as "a psychological state in which virtual social actors are experienced as actual social actors in either sensory or non-sensory ways" (p. 37). In other words, social presence is concerned with one's experiences with other humans or artificial social actors that manifest humanness connected by technology (Lee, 2004).

The extant body of research documents that social presence can be understood through various dimensions, such as psychological involvement and copresence (e.g., Biocca et al., 2003; Kelly & Westerman, 2016). *Psychological involvement* refers to a feeling of "psychologically" sensing another entity, and it is originated from the initial social presence research by Short et al. (1976). This aspect of social presence is about the connection with another agent in a mediated environment and/or the mutual awareness of another agent. *Copresence* refers to the perception of being "physically" with another entity in the same place (Biocca et al., 2003). In this regard, technology users may forget they are in a mediated context and feel as if their mediated experience was in a shared, physical space. In other words, people feel like they are with another social entity in the same space although they are physically apart from each other.

These two aspects are both conceptually related but distinct (J. Kim et al., 2016; Westerman et al., 2018). To clarify, although feeling physically present might be assumed as part of being psychologically present, it does not mean they always occur simultaneously. A person may experience a strong psychological sense of social richness, but it does not guarantee or indicate that they feel the sense of physically being together in the same space. However, when people feel each other's social nature, although they are aware of the physical distance from each other, they may experience the feeling of being together, although this feeling may only be momentary.

In an online learning environment, social presence plays a critical role in fostering positive learning experiences. A significant body of research documents that social presence positively influences student learning experiences, such as class participation and motivation to learn (Mazzolini & Maddison, 2007), course and instructor satisfaction (Akyol & Garrison, 2008; Strong et al., 2012), and learning outcomes (Hostetter, 2013; Joksimović et al., 2015; Kang & Im, 2013; J. Kim et al., 2016). Overall, these findings are well demonstrated in a meta-analysis that social presence is positively related to students' satisfaction with online learning experiences and perceived learning outcomes (Richardson et al., 2017).

In all, the fact that students experience social presence of their teachers in online courses, even though their teachers are not physically present, implies that students still develop perceptions about their teachers based on available cues online. During this experience, students may realize that they can still engage in meaningful learning in an online environment without a face-to-face interaction. In this regard, the present study questions whether this positively experienced social presence of a human teacher would translate into developing certain expectations or anticipations about a teacher that is a machine who serves a similar role as a human teacher.

Developing Expectations About Machine Teachers

Machine Teacher

The extant research indicates that a *machine teacher* is "a technology that plays a meaningful role during an interaction with humans in helping them engage in affective, cognitive, and behavioral learning through various ways" (J. Kim et al., 2020, p. 1904). Machine teachers can appear in diverse forms (J. Kim et al., 2020). For example, social robots (e.g., Pepper, NAO) and telepresence robots can be effectively used in a face-to-face pedagogical context.

Virtual agents (e.g., chatbots, software agents) can engage in interactions through text or voice in an online learning environment.

Although the idea of machine teachers is relatively new, scholars note a significant potential of machine teachers in an online learning environment for higher education. J. Kim et al. (2020) explored how undergraduate students perceive an AI teaching assistant in online education. They found that perceived usefulness of and perceived ease of communicating with an AI teaching assistant predict positive attitudes toward an AI teaching assistant. Then, the positive attitudes eventually predict intentions to take an AI teaching assistant-based online course. J. Kim et al. (2021) further examined the role of communication styles of an AI instructor in online education and found that undergraduate students experience more positive perceptions (e.g., attitudes) about a relational AI instructor than a functional, task-oriented AI instructor. Importantly, J. Kim et al. (2021) highlight the important role of social presence. That is, a relational AI instructor fosters stronger social presence of the AI instructor than a functional AI instructor; then, heightened social presence of the AI instructor facilitates positive perceptions about the AI instructor. Students' intentions to take an AI-based education depends on the availability of such courses, which should be first adopted by universities. Considering the potential of such availability in the future, J. Kim et al.'s (2020, 2021) findings suggest that developing positive perceptions about a machine agent is a key for the successful adoption of an AI-based education.

Human Responses to Machines

There is a growing interest in the understanding of how humans respond to various forms of technology. Currently, there are a few perspectives that suggest humans treat technology in a similar way as they treat a human. In particular, the computers are social actors (CASA) paradigm states that people mindlessly apply social scripts to interactions with computers, and it is due to people focusing on social cues and failing to focus on asocial characteristics (Nass & Moon, 2000). In doing so, these individuals interact with computers in a similar manner to how they interact with other people. Over the years, the CASA paradigm has been examined in diverse technologies, such as robots (Fischer, 2011), chatbots (C. Edwards et al., 2014), exergames (J. Kim & Timmerman, 2018), and AI (C. Edwards et al., 2019). Generally, findings indicate that humans are mindlessly responding to technologies as they would in their interpersonal interactions with other humans. As newer and more technologies develop, Gambino et al. (2020) proposed an extension of the CASA paradigm. The extension of the CASA paradigm suggests that humans may develop and apply human-media social scripts when interacting with machines, rather than mindlessly applying human-human social scripts.

In another line of research, Spence et al. (2014) suggested the idea of "human-to-human interaction script" (p. 277). The core of the human-to-human interaction script is concerned with differences in humans' expectations when interacting with another human or a machine, with a preference for a human. Supporting the aforementioned argument, a series of empirical studies (e.g., C. Edwards et al., 2016; Spence et al., 2014) documents that people experience less uncertainty and stronger social presence when they anticipate interacting with another human compared to when anticipating an interaction with a robot. Noticeably, after an actual interaction with a robot, people experience less uncertainty and

greater social presence about the robot, compared to an initial expectation that was formed prior to the actual interaction (A. Edwards et al., 2019).

Overall, CASA (Reeves & Nass, 1996), an extension of CASA (Gambino et al., 2020), and the human-to-human interaction script (Spence et al., 2014) help us understand how humans perceive machine agents. Although not explicitly discussed, these perspectives may be a foundation for the idea that previous experiences with a human may contribute to developing certain expectations about a machine or interactions with a machine in a situation where both the human and the machine serve the same or similar roles in the same or similar context. However, there is lack of evidence in the extant literature that supports this conjecture.

Thus, this study investigates the above-mentioned inquiry in the context of machine teachers in online education. Of multiple roles that machine teachers can serve (e.g., primary instructor, teaching assistant), we focus on an AI teaching assistant. Specifically, we examine whether positively experienced social presence of a human teacher contribute to developing positive attitudes toward an AI teaching assistant. Then, we further examine whether the positive attitudes toward an AI teaching assistant would eventually influence intentions to take an AI teaching assistant-based online course when it becomes available.

> **H1a–b:** Social presence, particularly (a) social presence as psychological involvement and (b) social presence as copresence, of human teachers experienced in online courses influences the way students develop attitudes toward an AI teaching assistant in online courses.

> **H2:** Attitudes toward an AI teaching assistant predict intentions to take an AI teaching assistant-based online course.

Methods

Participants

For this study, undergraduate students from communication classes at a large public university in the US were recruited. In order to identify eligible participants and to ensure the quality of the data, a few steps were taken. First, we removed responses recorded from any attempt beyond one's first-time participation. Second, we removed responses from participants who failed an attention check, which occurred in the middle of the survey.

After completing the screening steps, the final sample consisted of 294 undergraduate students who have online course experiences. The sample included more females ($n = 188$: 63.9%) than males ($n = 106$: 36.1%). The average age of participants was 21.55 years ($SD = 4.29$). The sample consisted of White/Caucasian ($n = 153$: 52%), Latino/a/x or Hispanic ($n = 67$: 22.8%), Black/African American ($n = 49$: 16.7%), and other ethnic groups ($n = 25$: 8.5%).

Procedure

Data were collected using an online survey tool. Upon the university's IRB approval, a recruitment message was distributed to potential participants. Upon clicking on the survey

link, participants were asked to read the informed consent. Then they proceeded to the main page of the survey.

The survey consisted of three major sections. The first section included participants' pre-existing attitudes toward new technologies and previous experiences with online courses. Specifically, participants were asked to identify one online course they most recently completed and provide the name of the course. This was necessary to help them think about their recent online experience. Then they were asked to answer questions based on their experiences of the particular course that they indicated.

The next section included the study material. In this section, participants were asked to read a short story about an AI teaching assistant. The article primarily explained the tasks that the AI teaching assistant performed in an online learning management site, such as responding to students' questions about the course and assignments. The article was written in a way that does not lead readers to create certain perceptions or expectations about an AI teaching assistant (see Appendix A). To ensure that participants read the article before proceeding to the next page, a timer was set to prevent them from skipping the task.

The last section focused on assessing participants' responses about the article they read. In particular, this section assessed how students perceived the AI teaching assistant. At the end of the survey, demographic questions were asked. All participants received extra credit and confidentiality was guaranteed.

Measures

At the start of the survey, participants' perceptions about social presence of a teacher and preexisting attitudes toward new technologies were assessed. *Social presence as psychological involvement* (α = .92) was evaluated with eight items (e.g., When I was taking the online class, I felt like my teacher was . . . "remote—immediate," "unsociable—sociable," and "impersonal—personal"). Items were adopted from the extant literature (Lombard et al., 2009; Short et al., 1976) and slightly modified for the study context. Specifically, the phrase, "when I was taking the online class, I felt like my teacher was . . ." was added to provide participants with the study's context for their responses. Responses were obtained on a 7-point semantic differential scale.

Social presence as copresence (α = .95) was assessed with four items (e.g., When I was taking that online class . . . "I felt like my teacher was with me" and "I felt like my teacher was interacting with me in the same space"). Items were adopted from Lee et al. (2006) and slightly modified for the study context. The original items were focused on an interaction with a social robot, AIBO. To make the items fit in the study context, interaction with AIBO-related phrases were replaced with taking an online class/perceptions about a teacher. Responses were recorded on a 7-point scale (1 = *Strongly Disagree*, 7 = *Strongly Agree*).

Preexisting attitudes toward new technologies (α = .89) were evaluated with three items (e.g., "How comfortable would you be with new technologies—e.g., robots, AI—taking interpretive roles (e.g., editorial writers, newspaper reporters, novelists)" and " . . . taking personal roles (e.g., colleagues, bosses)"). Items were adopted from Nass and colleagues (1995). Responses were obtained on a 6-point scale (1 = *Very uncomfortable*, 6 = *Very comfortable*).

After learning about an AI teaching assistant described in the study material (see Appendix A), participants' responses were evaluated. *Attitudes toward an AI teaching assistant* (α = .95) was assessed with adjectives on a 7-point semantic differential scale. The measure was adopted from Davis (1993). *Intentions to take an AI teaching assistant-based course* (α = .95) was assessed with three items (e.g., "If an AI teaching assistant-based online class is available, I intend to take the class," " . . . I would consider taking the class"). The measure was slightly modified from the extant research, which focused on the adoption of autonomous vehicles (Choi & Ji, 2015) to fit with the present study's context, the adoption of an AI teaching assistant-based course. Responses were recorded on a 7-point Likert-type scale (1 = *Strongly Disagree*, 7 = *Strongly Agree*).

Results

First, correlations between all tested study variables were assessed (see Table 1). Then, we conducted a path analysis, which is a particular type of structural equation modeling, to test the proposed hypotheses. A path analysis was used because it has the advantages of evaluating and presenting the comparative strengths of different relationships in the model[1] (Lleras, 2005). Mplus 7 was used to test the model (Muthén & Muthén, 2015). Maximum likelihood estimation was used when the analyses were conducted. According to Hu and Bentler (1999), a model has good fit when the Chi-square test is non-significant, *CFI* > .95, *RMSEA* < .06, and *SRMR* < .08. Based on the guidelines of Kline (2005) and Schreiber et al. (2006), after controlling for participants' age, sex, and preexisting attitudes toward new technologies, the result indicated that the model has a reasonable goodness of fit for the data, *AIC* = 1819.93, *BIC* = 1867.77, X^2 (2, N = 293) = 6.92, *p* = .03, *CFI* = .98, *TLI* = .91,

TABLE 1 Zero-Order Correlations Among Study Variables						
	1	**2**	**3**	**4**	**5**	**6**
1. Age	–					
2. Attitudes toward new technologies	−.12*	–				
3. Social presence as psychological involvement	.10	−.07	–			
4. Social presence as copresence	.08	.10	.65**	–		
5. Attitudes toward an AI teaching assistant	−.13*	.27**	.18**	.09	–	
6. Intentions to take an AI teaching assistant-based course	−.09	.26**	.15*	.16**	.77**	–
M	21.55	3.01	4.97	4.08	4.42	4.24
SD	4.29	1.38	1.20	1.63	1.43	1.50

*p < .05, **p < .01

---

1. Measurement models were not used in this study because the scales used in this study were already established and reliable in the previous studies. Thus, variables were treated as manifest variables in the model.

RMSEA = .09, *SRMR* = .01. Although the Chi-square test was significant and RMSEA was larger than .06, the model fit can still be considered acceptable for the following reasons: (a) the Chi-square test is sensitive to the sample size, (b) the value of RMSEA that is smaller than .10 can still be considered as a mediocre fit (Hooper et al., 2008; MacCallum et al., 1996), and (c) a set of fit indices determine the goodness of fit as a whole (Kline, 2005).

H1a–b predicted that social presence of a human teacher experienced in an online course, particularly social presence as psychological involvement (H1a) and social presence as copresence (H1b), would influence the way that students develop attitudes toward an AI teaching assistant in an online course. Regarding H1a, social presence as psychological involvement significantly and positively predicted attitudes toward an AI teaching assistant, *B* = .31, *SE* = .04, *p* < .001. With regard to H1b, social presence as copresence did not significantly predict attitudes toward an AI teaching assistant, *B* = –.06, *SE* = .06, *p* = .32. H1a was supported, but H1b was not.

Next, H2 proposed that attitudes toward an AI teaching assistant would predict intentions to take an AI teaching assistant-based online course. As predicted, there was a statistically significant and positive association between the two variables, *B* = .79, *SE* = .04, *p* < .001. H2 was supported. See Figure 1.

After the primary hypotheses testing, we conducted an additional analysis to understand which type of social presence students felt more strongly in their online learning experiences. Results from a paired *t*-test indicated that students reported stronger psychological involvement (*M* = 4.97, *SD* = 1.20) than copresence (*M* = 4.08, *SD* = 1.63). The difference was statistically significant, *t*(293) = 12.25, *p* < .001.

FIGURE 1   Final Model
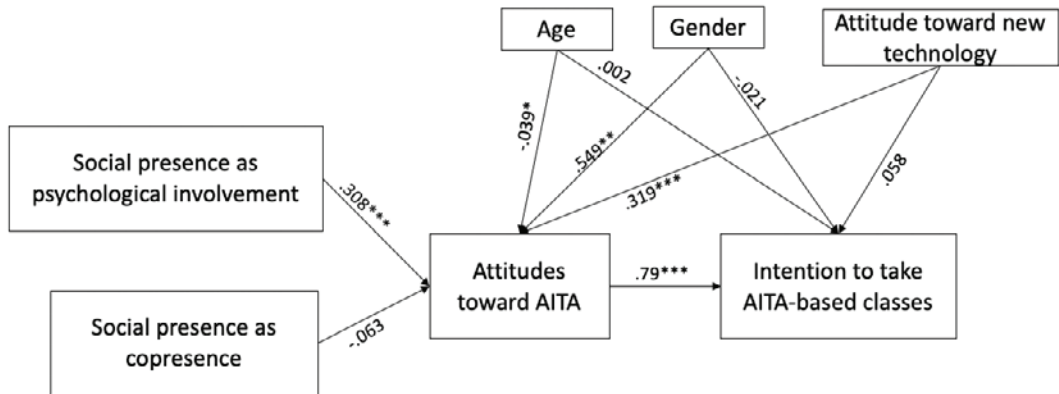
Note 1. X^2 (2) = 6.915; *p* = .03; CFI = .984; RMSEA = .092; SRMR = .014
Note 2. * *p* < .05, ** *p* < .01, *** *p* < .001
Note 3. AITA refers to AI teaching assistant

Discussion

We examined whether positively experienced social presence of a human teacher in an online course would influence the way that students develop certain expectations about a machine teacher, specifically an AI teaching assistant. Overall, findings indicate that social presence of a human teacher plays an important role in developing positive attitudes toward an AI teaching assistant, which consequently leads to intentions to adopt an AI teaching assistant-based online education when it becomes available. The following sections discuss implications and contributions of the study's key findings and future research directions based on limitations identified in this investigation.

Primary Findings, Implications, and Contributions

First, we found that positively experienced social presence, particularly social presence as psychological involvement, of a human teacher influence developing positive attitudes toward an AI teaching assistant. From a broad perspective, this finding confirms the extant research that documents an important role of social presence in fostering positive experiences in an online learning context (e.g., J. Kim et al., 2016; Sellnow & Kaufmann, 2018; Song, Kim, & Park, 2019). However, to the best of our knowledge, no research has shown that positively experienced social presence of a human teacher could influence developing positive attitudes about a machine teacher, an encounter that may occur in the future. In this regard, this finding makes a notable contribution to the existing literature.

Although social presence as psychological involvement is found to be significant, we did not find a significant role of social presence as copresence. This suggests that dimensions of social presence may work differently or may serve different roles. In fact, this argument is supported by the research indicating that social presence as psychological involvement and as copresence are related but distinct (J. Kim et al., 2016; Westerman et al., 2018). Another possible explanation is that students may not feel copresence as strongly as they feel psychological involvement in an online class environment, which may have consequently influenced the way copresence could predict attitudes toward an AI teaching assistant. In fact, this tendency appeared in the present study, as students reported stronger social presence as psychological involvement than social presence as copresence.

Although social presence as copresence did not produce a significant role, it is still important to consider it along with psychological involvement. In fact, the existing literature highlights that when diverse dimensions of social presence are considered together, it increases the predictability of learning outcomes in online education (J. Kim et al., 2016). Thus, the pattern found in the study that copresence is weaker compared to psychological involvement signals a need to find ways to facilitate a feeling of shared space. The extant literature reports that perceived similarity of the interaction partner fosters a strong sense of copresence (Song, Kim, & Choi, 2019) and extroverted individuals tend to feel strong copresence (J. Kim et al., 2018). These findings suggest that teachers are encouraged to find a way to create an environment where students perceive some similarity with teachers.

The finding that positively experienced social presence of a human teacher helps develop positive attitudes toward an AI teaching assistant signals a need to find ways to foster social presence of human teachers as the new era of an AI-based online education has started (J. Kim, 2021). Lessons can be drawn from the extant literature. A good deal

of research documents various factors that facilitate social presence of others in an online environment, such as self-disclosure (J. Kim & Song, 2016; J. Kim & Yang, 2019; Song, Kim, & Park, 2019), perceived similarity (Song, Kim, & Choi, 2019), and supportive feedback (J. Kim & Timmerman, 2018). More germane to the study's context, an instructor's self-disclosure is an important way to foster social presence (Song, Kim, & Park, 2019). In preparation for the new era of an AI-based online education, it is important to further investigate how to foster social presence of a human teacher in an online learning environment.

Further, attitudes toward an AI teaching assistant significantly predict one's intentions to take an AI teaching assistant-based course when it becomes available. In fact, the relationship between attitudes of a new technology and intentions to adopt that technology is well supported in various adoption models, such as the Technology Acceptance Model (TAM; Davis, 1989), the Theory of Planned Behavior (TSP; Ajzen, 1991), and the Value-based Adoption Model (VAM; H.-W. Kim et al., 2007). Also, empirical research findings support the relationship between attitudes and intentions (e.g., J. Kim et al., 2020; Sohn & Kwon, 2020). Sohn and Kwan found that attitudes related to AI-based intelligence products significantly predict one's intentions to adopt that AI-based intelligence product. More germane to the current investigation, J. Kim et al. (2020) found that attitudes toward AI teaching assistants predict intentions to adopt an AI teaching assistant-based online education. Thus, the present study's finding reemphasizes the important link between attitudes and intentions when considering the adoption of a new technology.

Overall, the study's findings contribute to new knowledge concerning how humans perceive and develop expectations about machines. The CASA paradigm suggests that humans treat machines in a similar manner to how they treat other humans, and this is often noted as an automatic process where humans assign social conventions to machines (Reeves & Nass, 1996). While CASA's perspective provides foundational understanding of how humans treat machines, it does not directly or explicitly state that experiences with humans can influence developing expectations or perceptions about a machine. In this regard, the study's finding that past experiences with humans can affect expectations about machines in similar roles (e.g., teacher-teacher) is a significant addition to the extant knowledge and suggests venues for future research.

More broadly, the study's finding has meaningful implications for human-machine communication (HMC). Guzman (2018) conceptualizes HMC as the creation of meanings between humans and machines. Research on HMC encompasses the scholarship of human-computer interaction, human-robot interaction, and human-agent interaction. It also investigates the sociocultural and critical aspects of these emerging technologies (Guzman, 2018). One advantage of incorporating machine agents into education is that machines are efficient at presenting pre-scripted messages to learners. Humans can carefully craft and edit educational scripts and utilize machine agents to deliver the messages. Although the idea of machine teachers is new to many educators and educational institutions, the use of machine agents in education has the potential to grow and make meaningful contributions to the study of HMC.

It is not clear yet when a machine teacher or AI-based education will be readily available in higher education. However, in preparation for the coming future, it is important to consider educating and training human teachers for effective use of a machine teacher or teaching assistant in an AI-based education. One should consider training for teachers

or teacher certification programs that teach instructors how to use AI effectively and help them realize the value AI teaching assistants can bring to the teaching and learning experience (C. Edwards et al., 2018).

Limitations and Future Research Directions

Although we report meaningful findings in this investigation, we acknowledge limitations that should be further examined in future research. First, the scope of the study is limited to assessing the intentions to take an AI teaching assistant-based online course, rather than the actual behavior. Although extant theoretical perspectives, such as the TAM (Davis, 1989), indicate a strong relationship between intentions and actual adoption, there may exist a gap between the two variables. When an AI-based education becomes readily available, it is necessary to examine how students' attitudes translate to adoption.

Second, participants in this study learned about AI teaching assistants by reading an article rather than directly interacting with it. It is likely that a direct interaction with an AI teaching assistant could elicit different responses among the participants. When the use of AI teaching assistants become readily available in the education system, future research should investigate whether students would respond differently to an AI teaching assistant after a direct interaction.

Lastly, we did not specify the nature or type of courses (e.g., topic, structure) when assessing students' intentions to take an AI teaching assistant-based online course. A seminar-based versus a lecture-based course may require AI teaching assistants to demonstrate different skills in assisting students in understanding course materials, completing worksheets or activities, or engaging in discussions. Teaching a social science versus a hard science course may also involve different means of course organizations that differ in terms of AI teaching assistants' responsibilities. Therefore, future research could examine whether the current findings can be generalized to all subject areas with various responsibilities of AI teaching assistants in online courses.

Conclusion

The present study examined whether previously experienced social presence of a human teacher in online courses would influence developing certain expectations about a machine teacher. Primary findings indicate that social presence as psychological involvement of a human teacher positively predicts attitudes toward an AI teaching assistant. Then, the attitudes predict intentions to take an AI teaching-based online course when it becomes available. This research is one of the first that demonstrates that prior experiences with a human can potentially influence one's expectations about a machine that performs similarly to the aforementioned human.

Because the adoption of machine teachers is arguably still in its infancy, pervasive use of them in mainstream online education may take years. However, technology-enhanced instruction is only growing and using machine teachers as assistants and perhaps also as primary instructors is likely to be as well. Thus, based on the present study's exploratory findings, future researchers are encouraged to further investigate this important area of research.

Author Biographies

Jihyun Kim is an Associate Professor in the Nicholson School of Communication and Media at the University of Central Florida. Broadly, her research is focused on the effects and implications of new media/communication technologies for meaningful outcomes (e.g., education, health). Her research examines the role of technology not only as a tool but also as a digital agent, which is centered on human-machine communication. She is particularly interested in people's perceptions about AI and social and psychological impacts of AI from communication perspectives. Additionally, her research is primarily driven by the theoretical notion of presence, especially social presence.

   ⓘ   https://orcid.org/0000-0003-2476-610X

Kelly Merrill Jr. is a doctoral candidate in the School of Communication at The Ohio State University. His primary research interests are at the intersection of communication technology and health communication. In particular, he is interested in the prosocial uses of communication technology, specifically focusing on promoting physical, mental, and social health and well-being across various communication channels. He is also interested in responses to discrimination and addressing health disparities among marginalized and stigmatized individuals.

   ⓘ   https://orcid.org/0000-0002-1221-3789

Kun Xu is an Assistant Professor at the Department of Media Production, Management, and Technology at the University of Florida. Kun Xu's research focuses on the intersection of human-robot interaction, human-computer interaction, and psychological processing of media. His work investigates how people perceive, evaluate, and respond to technologies such as social robots, computer agents, and virtual assistants. He also examines how people use virtual and augmented reality (VR/AR) technologies to make sense of spaces and maintain social relationships. His research centers on the role of social cues and social presence in both physical and virtual communication contexts.

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

Deanna D. Sellnow is *Luminary* Professor of Strategic Communication in the Nicholson School of Communication and Media at the University of Central Florida. She conducts research on strategic instructional communication in a variety of contexts (e.g., face-to-face and technology-enhanced classrooms, risk/crisis, health, agricultural biotechnology, and biosecurity). She has conducted funded research for the USGS, USDA, DHS, and CDC, and has presented her research across the United States and in many countries around the world.

   ⓘ   https://orcid.org/0000-0003-4668-4359

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

Akyol, Z., & Garrison, D. R. (2008). The development of a community of inquiry over time in an online course: Understanding the progression and integration of social, cognitive and teaching presence. *Journal of Asynchronous Learning Networks*, *12*, 3–22.

Al Ghamdi, A., Samarji, A., & Watt, A. (2016). Essential considerations in distance education in KSA: Teacher immediacy in a virtual teaching and learning environment. *International Journal of Information and Education Technology, 6,* 17–22. https://doi.org/10.7763/IJIET.2016.V6.651

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

Choi, J. K., & Ji, Y. G. (2015). Investigating the importance of trust on adopting an autonomous vehicle. *International Journal of Human-Computer Interaction*, *31*(10), 692–702. https://doi.org/10.1080/10447318.2015.1070549

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 319–340. https://doi.org/10.2307/249008

Davis, F. D. (1993). User acceptance of information technology: System characteristics, user perceptions and behavioral impacts. *International Journal of Man-Machine Studies, 38*(3), 475–487. https://doi.org/10.1006/imms.1993.1022

Edwards, A., & Edwards, C. (2017). The machines are coming: Future directions in instructional communication research. *Communication Education*, *66*(4), 487–488. https://doi.org/10.1080/03634523.2017.1349915

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, C., Edwards, A., Spence, P. R., & Lin, X. (2018). I, teacher: Using artificial intelligence and social robots in communication and instruction. *Communication Education, 67,* 473–480. https://doi.org/10.1080/03634523.2018.1502459

Edwards, C., Edwards, A., Spence, P. R., & Shelton, A. K. (2014). Is that a bot running the social media feed? Testing the differences in perceptions of communication quality for a human agent and a bot agent on Twitter. *Computers in Human Behavior*, *33*, 372–376. https://doi.org/10.1016/j.chb.2013.08.013

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

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

Fischer, K. (2011). Interpersonal variation in understanding robots as social actors. *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 53–60. https://doi.org/10.1145/1957656.1957672

Gambino, A., Fox, J., & Ratan, R. A. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication*, *1*(1), 5. https://doi.org/10.30658/hmc.1.5

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

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods*, *6*(1), 53–60.

Hostetter, C. (2013). Community matters: Social presence and learning outcomes. *Journal of the Scholarship of Teaching and Learning*, *13*(1), 77–86.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E., & Hatala, M. (2015). Social presence in online discussions as a process predictor of academic performance. *Journal of Computer Assisted Learning*, *31*(6), 638–654. https://doi.org/10.1111/jcal.12107

Kang, M., & Im, T. (2013). Factors of learner–instructor interaction which predict perceived learning outcomes in online learning environment. *Journal of Computer Assisted Learning*, *29*(3), 292–301. https://doi.org/10.1111/jcal.12005

Kaufmann, R., Sellnow, D. D., & Frisby, B. N. (2016). The development and validation of the Online Learning Climate Scale (OLCS). *Communication Education, 65,* 307–321. https://doi.org/10.1080/03634523.2015.1101778

Kelly, S., & Westerman, D. K. (2016). New technologies and distributed learning systems. In P. L. Witt (Ed.), *Handbooks of communication science: Communication and learning* (Vol. 16, pp. 455–480). DeGruyter Mouton.

Kim, H.-W., Chan, H. C., & Gupta, S. (2007). Value-based adoption of mobile internet: An empirical investigation. *Decision Support Systems*, *43*(1), 111–126. https://doi.org/10.1016/j.dss.2005.05.009

Kim, J. (2021). A new era of education: Incorporating machine teachers into education. *Journal of Communication Pedagogy*, 4, 121–122. https://doi.org/10.31446/JCP.2021.1.11

Kim, J., Merrill Jr., K., Xu, K., & Sellnow, D. D. (2020). My teacher is a machine: Understanding students' perceptions of AI teaching assistants in online education. *International Journal of Human–Computer Interaction, 36*(20), 1902–1911. https://doi.org/10.1080/10447318.2020.1801227

Kim, J., Merrill Jr., K., Xu, K., & Sellnow, D. D. (2021). I like my relational machine teacher: An AI instructor's communication styles and social presence in online education. *International Journal of Human-Computer Interaction, 37*(18), 1760–1770. https://doi.org/10.1080/10447318.2021.1908671

Kim, J., & Song, H. (2016). Celebrity's self-disclosure on Twitter and parasocial relationships: A mediating role of social presence. *Computers in Human Behavior*, *62*, 570–577. https://doi.org/10.1016/j.chb.2016.03.083

Kim, J., Song, H., & Lee, S. (2018). Individual differences in social TV viewing experiences: A mediating and moderating role of social presence. *Mass Communication and Society, 21,* 50–70. https://doi.org/10.1080/15205436.2017.1350715

Kim, J., Song, H., & Luo, W. (2016). Broadening the understanding of social presence: Implications and contributions to the mediated communication and online education. *Computers in Human Behavior*, *65*, 672–679. https://doi.org/10.1016/j.chb.2016.07.009

Kim, J., & Timmerman, C. E. (2018). Effects of supportive feedback messages on exergame experiences. *Journal of Media Psychology*, *30*(1), 29–40. https://doi.org/10.1027/1864-1105/a000175

Kim, J., & Yang, H. (2019). How does a radio host's testimonial influence media experiences? The indirect effect of social presence. *Journal of Radio & Audio Media*, *26*(2), 336–350. https://doi.org/10.1080/19376529.2018.1509217

Kline, R. B. (2005). *Principles and practice of structural equation modeling.* The Guilford Press.

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

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

Limperos, A. M., Buckner, M. M., Kaufmann, R., & Frisby, B. N. (2015). Online teaching and technological affordances: An experimental investigation into the impact of modality and clarity on perceived and actual learning. *Computers and Education, 83,* 1–9. https://doi.org/10.1016/j.compedu.2014.12.015

Lleras, C. (2005). Path analysis. In K. Kempf-Leonard (Ed.), *Encyclopedia of social measurement* (pp. 25–30). Elsevier. https://doi.org/10.1016/B0-12-369398-5/00483-7

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

Lombard, M., Ditton, T. B., & Weinstein, L. (2009, November*).* Measuring presence: The Temple Presence Inventory. *Proceedings of the International Society for Presence Research Annual Conference,* Los Angeles, CA.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149.

Mazzolini, M., & Maddison, S. (2007). When to jump in: The role of the instructor in online discussion forums. *Computers & Education*, *49*(2), 193–213. https://doi.org/10.1016/j.compedu.2005.06.011

Muthén, L. K., & Muthén, B. O. (2015). *Mplus user's guide 7th edition.* Muthén & Muthén.

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

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

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.

Richardson, J. C., Maeda, Y., Lv, J., & Caskurlu, S. (2017). Social presence in relation to students' satisfaction and learning in the online environment: A meta-analysis. *Computers in Human Behavior*, *71*, 402–417. https://doi.org/10.1016/j.chb.2017.02.001

Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research, 99*(6), 323–338. https://doi.org/10.3200/JOER.99.6.323-338

Sellnow, D. D., & Kaufmann, R. (2018). Instructional communication and the online learning environment: Then, now, and next. In M. L. Houser & A. M. Hosek (Eds.), *Handbook of instructional communication: Rhetorical and relational perspectives* (2nd ed, pp. 195–206). Routledge.

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

Sohn, K., & Kwon, O. (2020). Technology acceptance theories and factors influencing artificial intelligence-based intelligent products. *Telematics and Informatics*, *47*, 101324. https://doi.org/10.1016/j.tele.2019.101324

Song, H., Kim, J., & Choi, Y. (2019). Social TV viewing: The effect of virtual co-viewers and the role of social presence. *Journal of Digital Contents Society, 20*(8), 1543–1552. http://dx.doi.org/10.9728/dcs.2019.20.8.1543

Song, H., Kim, J., & Park, N. (2019). I know my professor: Teacher self-disclosure in online education and a mediating role of social presence. *International Journal of Human–Computer Interaction*, *35*(6), 448–455. https://doi.org/10.1080/10447318.2018.1455126

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*(3), 272–280. https://doi.org/10.1080/08824096.2014.924337

Strong, R., Irby, T. L., Wynn, J. T., & McClure, M. M. (2012). Investigating students' satisfaction with eLearning courses: The effect of learning environment and social presence. *Journal of Agricultural Education*, *53*(3). https://doi.org/10.5032/jae.2012.03098

Westerman, D., Edwards, A., Edwards, C., & Spence, P. R. (2018, May). Social robots and social presence: Interpersonally communicating with robots. In *Human-Machine Communication Preconference at the International Communication Association Conference*, Prague, Czech Republic.

Appendix A

Here's a sample interaction between the chatbot and a student. (Courtesy Ashok Goel)

To help with his class this spring, a Georgia Tech professor hired Jill Watson, a teaching assistant unlike any other in the world. Throughout the semester, she answered questions online for students, relieving the professor's overworked teaching staff.

But, in fact, Jill Watson was an artificial intelligence bot.

Ashok Goel, a computer science professor, did not reveal Watson's true identity to students until after they'd turned in their final exams.

Students were amazed. "I feel like I am part of history because of Jill and this class!" wrote one in the class's online forum. "Just when I wanted to nominate Jill Watson as an outstanding TA in the CIOS survey!" said another.

Now Goel is forming a business to bring the chatbot to the wider world of education. While he doesn't foresee the chatbot replacing teaching assistants or professors, he expects the chatbot's question-answering abilities to be an invaluable asset for massive online open courses, where students often drop out and generally don't receive the chance to engage with a human instructor. With more human-like interaction, Goel expects online learning could become more appealing to students and lead to better educational outcomes.

"To me this is a grand challenge," Goel said. "Education is such a huge priority for the entire human race."

At the start of this semester Goel provided his students with a list of nine teaching assistants, including Jill, the automated question and answering service Goel developed with the help of some of his students and IBM.

Goel and his teaching assistants receive more than 10,000 questions a semester from students on the course's online forum. Sometimes the same questions are asked again and again. Last spring he began to wonder if he could automate the burden of answering so many repetitive questions.

As Goel looked for a technology that could help, he settled on IBM Watson, which he had used for several other projects. Watson, an artificial intelligence system, was designed to answer questions, so it seemed like a strong fit.

To train the system to answer questions correctly, Goel fed it forum posts from the class's previous semesters. This gave Jill an extensive background in common questions and how they should be answered.

Goel tested the system privately for months, having his teaching assistants examine whether Jill's answers were correct. Initially the system struggled with similar questions such as "Where can I find assignment two?" and "When is assignment two due?" Goel tweaked the software, adding more layers of decision-making to it. Eventually Jill reached the point where its answers were good enough.

"I cannot create chaos in my classroom. Jill had to be almost perfect as a human TA or I am," Goel said.

The system is only allowed to answer questions if it calculates that it is 97 percent or more confident in its answer. Goel found that was the threshold at which he could guarantee the system was accurate.

There are many questions Jill can't handle. Those questions were reserved for human teaching assistants.

Goel plans to use Jill again in a class this fall, but will likely change its name so students have the challenge of guessing which teaching assistant isn't human.

"A really fun thing in this class has been once students knew about Jill they were so motivated, so engaged. I've never seen this kind of motivation and engagement," Goel said. "What a beautiful way of teaching artificial intelligence."

💬 27 Comments

Economy & business email alerts
Important breaking news emails on the issues around the economy and business.

E-mail address   Sign up

Exoskeletons and the Future of Work: Envisioning Power and Control in a Workforce Without Limits

Gavin Kirkwood[1] , J. Nan Wilkenfeld[1] , and Norah E. Dunbar[1]

1 Department of Communication, University of California Santa Barbara, Santa Barbara, CA, USA

Abstract

Exoskeletons are an emerging form of technology that combines the skills of both machines and humans to give wearers the ability to complete physically demanding tasks that would be too strenuous for most humans (Sarcos Corp, 2019). Exoskeleton adoption has the potential to both enhance and disrupt many aspects of work, including power dynamics in the workplace and the human-machine interactions that take place. Dyadic Power Theory (DPT) is a useful theory for exploring the impacts of exoskeleton adoption (Dunbar et al., 2016). In this conceptual paper, we extend DPT to relationships between humans and machines in organizations, as well as human-human communication where use of an exoskeleton has resulted in shifts of power.

Keywords: exoskeletons, human-robot communication, interpersonal power, Dyadic Power Theory

There is a growing movement in industry to combine the strength, precision, and performance of machines with the agility, intelligence, and creativity of humans through the use of wearable robots, among other technologies (de Looze et al., 2016; Kong et al., 2019). One of the largest sectors within the $130 million wearable robotic industry includes the development of exoskeleton suits for medical, military, or industrial settings (Demaitre, 2019). Generally, industrial exoskeletons are defined as "a wearable device used to support and assist the strength and mobility of the wearer" (Upasani et al., 2019, p. 2). Human-centered

CONTACT Norah E. Dunbar   • Department of Communication • University of California Santa Barbara • 4423 Social Science and Media Studies, Santa Barbara, CA 93106-4020 • ndunbar@ucsb.edu

design seeks to increase the productivity of users of new technologies while improving the user experience, increasing accessibility, and reducing discomfort and stress (Giacomin, 2014). But new technologies often increase stress for the users, especially when technology adoption results in a user's role being open to change or reinterpretation. New technologies often create new roles and shift the balance of power toward those who can more easily adopt the new technologies. Decades of research on technology adoption in organizations have shown that emerging technologies can reshape relationships between coworkers, create role reversals, and disrupt expertise (Barley, 1986; Beane, 2019). In this conceptual paper, we discuss the potential of industrial exoskeleton technologies to shape human-machine and human-human power relationships across a variety of industries and theorize how power dynamics might change in these settings. As Fortunati and Edwards (2020) explain, the power imbalance between humans and machines necessitates adaptation on the part of human actors and can cause frustration when robots and other machines constrain our interactions.

Power is the ability to influence and affect the behavior of others (Dunbar, 2015). Specifically, in this paper we examine a theory of interpersonal power, dyadic power theory (DPT; Dunbar, 2004; Dunbar et al., 2016), which is an interpersonal theory of power that explains the effects of power differences on the outcomes of interaction such as satisfaction. We use DPT to explain how new technologies affect organizational power relationships (using exoskeletons as a case study) in two ways. First, human power hierarchies are based on status and access to resources. Adding a scarce new technology into the workplace means that those with access to that technology and the knowledge about how to use it will have increased power even if there isn't a change to the formal organizational hierarchy. Second, humans often treat machines like coworkers and anthropomorphize their interactions with other technologies like avatars (e.g., Gambino et al., 2020; Nowak & Biocca, 2003) and we expect that exoskeleton users will do the same thing, supported by anecdotal evidence of early adopters. We therefore follow previous scholarship in acknowledging the interdependent and communicative relationships between humans and increasingly agentic machines that impact power dynamics (Banks & de Graaf, 2020; Guzman, 2018). While there might be many other theories of power that could be relevant here (see Dunbar, 2015 for a review of interpersonal theories) and media theories of power that may also be relevant (see Fortunati's 2014 discussion of media tools as sources of empowerment), we emphasize the interpersonal relationships in the workplace that are affected by the introduction of exoskeletons which is why we chose DPT as our theoretical focus.

What Are Exoskeletons?

Passive exoskeletons do not have a power source; instead, these devices rely on counterweights to collect energy from the wearer's own movements. Passive exoskeletons are primarily used to support healthy postures or prevent injury in work that requires repetitive tasks. An example of an upper-limb passive exoskeleton currently on the market is the *EksoVest* (Ekso Bionics, 2019). The EksoVest is designed for workers who engage in repetitive overhead movements that can strain the upper limb, shoulders, and upper back area. The EksoVest can provide full range of movement, is fully customizable to all heights and body types of workers, and can offer a lift assistance range of 5–15 pounds.

In contrast to passive exoskeletons that reduce fatigue, prevent injuries, and minimize the degeneration associated with repetitive strain, *active exoskeletons* can be used to dramatically augment human abilities or performance in physical taskings (Zaroug et al., 2019). Active exoskeletons are powered through actuators, such as electric motors, pneumatics, levers, hydraulics, or some combination of these components (McGowan, 2018). Active exoskeletons were previously developed for use in military settings, including Raytheon's XOS 2 powered armor suit which gave wearers the capability to lift 200 pounds while exerting little physical energy (Kopp, 2011). Emerging forms of active exoskeletons leverage the same capabilities as the Raytheon XOS powered suit but are designed for industrial contexts in the private sector (Kara, 2018). An example of a full-body active exoskeleton being developed for industry is the Guardian XO suit by Sarcos Corp (Sarcos Corp, 2019). Similar to the XOS 2, the Guardian XO will allow humans to lift up to 200 pounds with little energy exertion and will also allow full range of motion so that wearers can perform highly precise tasks with industry-specific equipment. Additionally, the XO contains around 125 onboard sensors with roughly three servers worth of computing power in order to capture and analyze the massive amounts of data being collected by the suit as it's being worn (Horaczek, 2020). Data currently being collected primarily consists of movement information; however, future designs will include more robust information such as operating environment and diagnostics.

Although active exoskeletons are not widely available for commercial purchase, glimpses of these new forms of wearable technologies demonstrate the potential of active exoskeletons to transform work practices across traditionally blue-collar industries such as shipping warehouses, construction sites, manufacturing plants or distribution centers, and also other settings such as hospitals. Unlike passive exoskeletons which are more like a harness or heavy backpack, active exoskeletons are more like robots and are likely to be anthropomorphized, as we discussed above. One important social implication of active exoskeletons is the way these technologies may impact power dynamics in human-machine and human-human interactions. Given the potential for active exoskeleton adoption to transform human-machine interactions across many organizations, researchers need to have theories that can be used to explore the power balances felt in human-exoskeleton interactions. We turn to a discussion of those theories next.

Exoskeletons and Power

Power is one of the most important aspects of all human interactions because it operates *under the surface*, affecting the communication choices we make even if we are not overtly aware of them (Dunbar, 2016). When a new technology is introduced into the workplace, it has the potential to shift the balance of power between members in an organization (Burkhardt & Brass, 1990). Power shifts can be especially salient when there is a gulf of expertise between novices and advanced users of the technology. For instance, in their ethnographic research in the medical industry, Barley (1986) found that the adoption of computerized tomography (CT) scanners resulted in role reversals between radiologists and technologists. In these role reversals, radiologists relied on technologists to help identify pathologies in CT scans because although technologists were not supposed to diagnose pathologies, they were the most skilled at reading the scans (Barley, 1986). More recent

research on robot adoption in the workplace has shown that emerging technologies continue to disrupt expertise in roles (Beane, 2019). In a case study on a cadre of beginner surgeons, Beane found that the new collaborative relationships with robots in surgery interrupted the normal training process for surgeons and required that they prematurely chose an area of specialization. Although these studies do not explicitly mention changing power dynamics amidst technology adaption, it is clear that in role reversals and changes in expertise that organizational members experience changes in relative (or informal) authority. The change in relative authority across these contexts showcases the need for researchers interested in technology adoption to more critically engage with how power dynamics change in these contexts.

Emergent technologies also can impact interpersonal dynamics or disrupt levels of autonomy between different stakeholder groups in organizations (Guzley et al., 2002). A useful theory for exploring these phenomena is DPT, which looks at the dyadic nature of power and emphasizes the relative perceived power of two actors in a relationship (Dunbar et al., 2016). DPT is an especially relevant theory when discussing the adoption of emergent technologies because it addresses how individuals perceive their own level of power as well as power balances across their relationships. In DPT, an actor's perception of their power is influenced by two key factors: authority and access to resources. We use the exoskeleton context to extend DPT to relationships between humans and machines in organizations, as well as human-human communication where use of an exoskeleton has resulted in shifts of power.

Although DPT has largely been used in interpersonal communication, its clear explication of power variables and scalable potential across different units of analysis make it useful for exploring exoskeleton adoption in the workplace. In the following sections we explicate DPT mechanisms, demonstrate potential impacts of exoskeletons in workplace human-machine interactions, and extend core DPT propositions to relationships in the exoskeleton context (see Table 1). In our revised propositions, we apply DPT to two units of analysis including power balances in the relationship between humans and active exoskeletons and power distribution across work teams that use active exoskeleton technology.

Power Definitions and Interactional Phases

From a DPT perspective, an actor's perception of their power and perception of power balances in their relationships is influenced by authority and access to resources (Dunbar et al., 2016). However, in DPT, perceptions of power and power itself are explicated differently. In DPT, power is conceptualized as an ability of an actor to influence behavior of another to achieve context-specific goals or outcomes (Dunbar, 2015). Dunbar explained, based on the work of Komter (1989), that across interactions there are three types of power: manifest power, latent power, and invisible power. In manifest power there are visible displays of power within an interaction such as open conflict, identifiable verbal behavior, or nonverbal behavioral cues that lead to desired goals or outcomes. Latent power operates in interactions when a less powerful person identifies the needs and desires of a more powerful person and accommodates in order to avoid conflict. Invisible power includes social or psychological mechanisms that manifest themselves in systematic power inequities such as gender norms or racial inequalities. Although DPT was created to bring understanding to

how humans use power to influence behavior and achieve certain outcomes in social inter
actions, new forms of technology also can influence human behavior or constrain human
agency (Boudreau & Robey, 2005; Huber, 1990; Jones, 1999). Therefore, we argue that man-
ifest, latent, or invisible forms of power can be present in interactions between humans and
exoskeletons as well as between exoskeleton wearers and non-wearers in the workplace.
These types of power are evident in the discussion that follows because whether or not
power is evident or operating *below the surface* is a result of the relationship between the
two interaction partners.

TABLE 1 Applying DPT Propositions to the Exoskeleton Context		
DPT proposition	**Revised DPT proposition**	**Explanation**
P1: Increases in relative authority will be related to increases in relative resources.	P1: Relative authority and access to resources will have a positive reciprocal relationship.	In the context of technology adoption, access to the exoskeleton will help wearers gain expertise with the suit which will increase their relative authority in the workplace.
P2: Increases in relative resources produce an increase in relative power.	P2: Increases in relative resources for humans or machines will produce an increase in relative power.	Exoskeletons that are imbued with enhanced capabilities for surveillance or workflow management will have increased power over wearers.
P3: Increases in relative authority produce an increase in relative power.	P3: Increases in relative authority for humans or machines will produce an increase in relative power.	Exoskeletons that imbued with relative authority to guide and influence wearer behavior will have increased power over wearers.
P4: The relationship between perceived power and control attempts is curvilinear.	No revision needed.	Exoskeleton wearers (at the individual, team, or department levels) perceive less power distance between each other and will engage in control attempts such as disciplining each other.
P5: Greater control attempts will lead to more control over outcomes.	P5: In human-machine relationships with high power discrepancies, control attempts by the more powerful actor are likely to succeed.	When a more powerful actor (whether human or exoskeleton) exercises a control attempt they are more likely to achieve their/its desired outcome in an interaction.
P6: As a partner's perception of their own power relative to their partner's power increases, counter-control attempts will increase as well.	No revision needed.	Exoskeleton wearers (either at the individual or team level) who engage in control attempts will likely be met with counter-control attempts.

DPT proposition	Revised DPT proposition	Explanation
P7: Counter-control attempts will hinder control over outcomes.	No revision needed.	Counter-control attempts mitigate the risk of one wearer or team from dominating their peers and may be needed to keep power balances in check.
P8: The relation between perceived relative power and relational satisfaction is curvilinear.	P8: Perceived power imbalances between humans and machines will have a negative effect on overall job satisfaction.	Perceived power imbalances, such as decreases in autonomy as a result of exoskeleton adoption, can lead to a variety of factors (such as increased stress) which will decrease overall job satisfaction.
P9: Control/counter-control attempts will be more frequent in high-high dyads than low-low dyads.	No revision needed.	In organizations where exoskeleton technology is considered valuable, workers who use exoskeletons will have higher power balances and are likely to engage in more control attempts of one another.
P10: Within generally power-balanced relationships, conflict will occur more often in domains over which both partners have high power than in domains in which only one partner has high power.	No revision needed.	Team members of equal status who have high power in the exoskeleton domain are more likely to engage in conflict over use of this technology.
P11: Dyads in cultures that stigmatize open conflict will display less conflict than dyads in cultures that do not stigmatize open conflict.	No revision needed.	While the corporate culture where the exoskeleton is deployed might be relevant, it is not explicitly discussed in this paper.
P12: Heterosexual couples in strongly patriarchal cultures will have less conflict than couples in less patriarchal cultures.	No revision needed.	Not relevant to exoskeleton context.
P13: Equal-power organizational dyads will display more control attempts than unequal-power dyads.	No revision needed.	Organizations that have equal access to exoskeletons are more likely to engage in control attempts such as industry regulations than organizations with unequal access to resources.

DPT proposition	Revised DPT proposition	Explanation
There is no P14 in DPT	P14: Teams with the agility to shift power domains between members will increase their effectiveness.	If teams are able to shift power based on situational demand, or expertise of team members then they will work more effectively.

Given that DPT is focused on power dynamics in dyadic social interactions, DPT propositions are separated into three distinct phases: pre-interactional phase, an interaction phase, and the post-interaction phase (Dunbar et al., 2016). DPT draws on the social exchange perspective and "predicts that pre-existing cultural, relational, and social factors and the resources that one has access to determine perceptions of one's own power that influences their behavioral tactics within social interactions" (Dunbar, 2015, p. 7). In the pre-interaction phase, the two pre-conditions of power that are likely to shape an interaction include authority and resources (Dunbar, 2015; Dunbar et al., 2016). The interaction phase and post-interaction phase are discussed in later sections.

Authority

Dunbar (2004) defines authority as norms regarding who *ought to* have control in a relationship. For example, whether to adopt exoskeleton technology in the workplace is likely a decision to be made by individuals who have a legitimized form of authority in an organization (such as CEOs, managers, or other individuals with formal decision-making power and spending authority). Authority is always in relation to the interaction partner (called *relative authority* in DPT terms) and can also afford the power to impact organizing processes as individuals with valuable skills, knowledge, or expertise can be influential in the workplace. An example could be that early adopters of exoskeleton technology have relative authority based on their expertise with the technology; they could use this authority to influence perceptions and coworker attitudes toward these technologies through manifest power. Additionally, increases in relative authority could lead to a hierarchical system in which users of the exoskeleton are seen as more valuable to the organization.

When considering human-machine interactions, scholars have long debated whether the human or the machine should have the authority to be the *leader* of the interaction (Draper et al., 1964 as cited in Kirkwood et al., 2021; Major & Shah, 2020). Should the human adapt to the exoskeleton or vice versa (or both)? This question conceptualizes technology as agent rather than tool in order to understand the influence on relationships, processes, and organizational structures and paves the way for researchers to apply human-human theoretical lenses to human-machine interactions (Gibbs et al., 2021). Previous research has found that people synchronize their behavior to machines despite the fact that these behaviors typically do not benefit their machine partners (Fujiwara et al., 2021). Although exoskeletons do not have artificial intelligence, humans may feel that they have to adapt to the machine when the machine does not follow their lead. This may reflect the common *media equation* effect which suggests that people unconsciously treat computers as social actors (Reeves & Nass, 1996).

In a recent unpublished pilot test with full-bodied, powered exoskeletons, users described working with the machine much like they would describe working with a human counterpart, " . . . *and, um, it would be really clunky to do that because that's—I guess not how the robot wants to walk*" *(P3).* Participants felt they were the recipient of actions, rather than the initiator at times, "*it felt like sometimes I'd be pulled or shoved in a direction that I, I didn't feel like I caused*" *(P8).* Additionally, many described battling with the machine for control over movement, "*if it's fighting you, uh, in an attempt to do something then that's usually a clear indication that it's not doing what it needs to do*" *(E4).* These statements provide preliminary (and anecdotal) evidence that people perceive these machines as having some level of agency and feel the need to understand and adapt to the exoskeleton.

Resources

Dunbar's (2004) original conception of DPT defined resources as anything that helps a partner satisfy needs or attain goals. In this way, an exoskeleton is a resource that helps humans complete tasks but they also have the power to disrupt *organizational membership* in nuanced ways. Boudreau and Robey (2005) argued that tensions between human agency and material agency have led to skewed perspectives that overemphasize either one's effect on the other. While some perspectives overemphasize human agency to enact technology in varied ways (even in ways contrary to how the technology was designed; Orlikowski, 2000), other perspectives overemphasize the ability of technology to determine or shape human behavior (Huber, 1990). In response to the polarization between human and material agency in sociotechnical relationships, Jones (1999) offered a dialectic and emergent perspective in which, "the particular trajectory of emergence is not wholly determined either by the intentions of the human actors or by the material properties of technology, but rather by the interplay of the two" (p. 297). Due to the expense associated with powered exoskeletons, if they are a scarce resource available to only a few employees, having access to exoskeletons is a resource in and of itself.

Resources, Diversity, and Inclusion

Active exoskeletons can offer every employee the same lifting or moving abilities, this may diversify the types of candidates that are well-suited for traditionally physically demanding work. Although a person with physical limitations might currently be excluded from physically demanding jobs, active exoskeleton technology may allow the person to perform tasks they previously were unable to, such as lifting heavy objects or squatting. The diversification of organizational membership in traditional blue-collar industries (such as auto manufacturing or shipping warehouses) has implications at the individual and organizational level. On the individual level, diversifying organizational membership can challenge traditional definitions of expertise and professional identity which may impact whether laborers who once took pride in their strength will continue to do so. By disrupting what types of expertise are valued in organizations, through invisible forms of power, exoskeleton technology can have a direct impact on the relative authority of workers; disruptions to relative authority can impact an employee's perceptions of their power in social interactions. On the organizational level, diversifying organizational membership can help create a more inclusive

environment which could lead to higher performance outcomes and better, more efficient problem-solving (Harrison & Klein, 2007; Jehn & Bezrukova, 2004).

Exoskeletons and Discriminatory Practices

Although active exoskeletons afford employers the opportunity to create a more diverse and inclusive workplace, the technology could also be used to discriminate against job candidates. The potential for an active exoskeleton to increase diversity and inclusion practices in organizations will be determined by organizational practices and how the suits are designed. Active exoskeletons are expensive to manufacture, and it is not clear which body types can be accommodated in the suits as they are currently created on a one-size-fits-all system (Zhang et al., 2017). For instance, while Sarcos Corp (2019) currently markets the enhanced physical capabilities afforded by the Guardian XO, the company does not provide information regarding the height requirements, weight limitations, or physical limitations that restrict who can wear the suit. While some suits have the capability to make minor sizing adjustments, humans vary widely in body shape and suits may not be designed to accommodate all body types. In addition to sizing, patterns of body movements vary from person to person. It is likely that individuals will have varying physical experiences with the exoskeleton with some finding it more challenging or more natural to embody than other users (Zhang et al., 2017). Knight and Baber (2005) emphasized that feelings of discomfort or pain are salient issues in wearable technology, and workers who have sustained workplace injuries or who have physical disabilities could be especially vulnerable to feelings of discomfort in the suit. Other industries have historically imposed weight or height requirements in order to exclude individuals from organizational membership (Murphy, 1998). In her study on flight attendant resistance, Murphy examined a case in which flight attendants challenged weight requirements and ultimately had the airline overturn those requirements. While the flight attendants in Murphy's study were able to create a more inclusive work environment, employees who are excluded from wearing an active exoskeleton may come across more obstacles when challenging their employer. These challenges can be exacerbated if it is expensive for organizations to modify the suits to make them adaptable or adjustable to all body types. Exoskeletons that cannot accommodate diverse body types may afford employers the power to exclude employees from roles which require exoskeleton usage; this could result in discriminatory practices within organizations. Thus, while exoskeletons are an important resource for workers in a variety of fields, they might operate to both enhance power equality by making certain physical tasks available to workers currently excluded from those opportunities but may also further exacerbate power inequality by highlighting physical differences that may lead to discrimination. Each individual organization should evaluate the effect that exoskeletons are providing to their employees based on their uses and availability.

Revised DPT Propositions

DPT originally had eight theoretical propositions (Dunbar, 2004). An additional five propositions were added in a subsequent revision and expansion of the theory (Dunbar et al., 2016). We explore how DPT can enlighten our understanding of the introduction of

exoskeleton technology to the workplace by systematically considering several of the theory's propositions in this context. Although not all the propositions are relevant to the exoskeleton context, we offer some discussion of the most relevant propositions and how they might apply in exoskeleton-human interactions. We hope readers will find this expanded theory useful when studying the integration of exoskeletons into workforces (see Table 1 for a summary of the original propositions and revisions). While this paper addresses theoretical questions, methodological issues in testing DPT are addressed elsewhere (Dunbar et al., 2016). The first three propositions relate to the pre-interaction phase in dyadic interactions.

Pre-Interactional Propositions

The first proposition of DPT (P1) is that increases in relative authority will be related to increases in relative resources. Although it is certainly the case that legitimized authority in the organizational context can lead to increases in greater resources, it is also possible for increased access to resources to increase a person's relative authority. Relative authority can be operationalized with a variety of context-specific variables. In the technology adoption context, expertise with new technologies will increase an individual's relative authority. Within the exoskeleton context, early wearers are likely to be sought out by other organizational members for information about the suit, knowledge of how to use the suit effectively, and how work tasks may need to be modified considering these new technologies. Given this logic, we revise P1 accordingly:

> **P1:** Relative authority and access to resources will have a positive reciprocal relationship.

In this revised proposition it is important to *reiterate the distinction between relative authority and legitimized authority.* We are not arguing that access to an active exoskeleton will lead to a formal increase in authority (such as a promotion to management, etc.). Rather, relative authority can be informal in nature such as coworkers considering an early adopter of active exoskeletons as an expert in the technology.

The second and third propositions of DPT indicate that increases in relative resources (P2) and relative authority (P3) will produce an increase in relative power (Dunbar et al., 2016). While P1 explicates the relationship between relative authority and resources, P2 and P3 explicate the direct relationship that relative authority and access to resources have on relative power in social interactions. We argue that these propositions hold true not only for humans but for active exoskeletons as well. Some organizations may imbue the exoskeleton with additional resources (in the form of technological capabilities) that allow the technology to surveil employees and influence employee behavior. New capabilities afforded by algorithmic management and RFID tags have exemplified cases in which employers use technology to closely monitor their employees, collect personalized data on how employees work, and use that information to terminate employees or influence employee behavior (Chan & Humphreys, 2018; Lupton, 2020). While exoskeleton manufacturers have not marketed surveillance capabilities in active exoskeletons, these suits contain hundreds of sensors that collect and process information about a wearer's movements (Islam & Bai, 2020). While this data is necessary to control the exoskeleton, there are certainly opportunities

for collecting and using this data to surveil and track employee movement and productivity. These forms of increased resources or capabilities of active exoskeletons are likely to increase the power these devices have in human-machine interactions. We revise P2 accordingly:

> **P2:** Increases in relative resources for humans or machines will produce an increase in relative power.

P2 is also relevant for human actors in dyadic interactions, as access to a resource such as an active exoskeleton will increase the relative power that the wearer has in interactions with non-wearers. It also increases the power of managers to surveil their employees and thus acts as a resource for them as well.

P3 directly addresses the positively associated relationship between relative authority and relative power. In addition to collecting personalized data on employee behavior and employee movement, active exoskeletons may also have a pedagogical element similar to other wearable technology such as Fitbits. Just as algorithms already manage employees in the rideshare context (e.g., Uber or Lyft; Rosenblat, 2019), it is possible that exoskeletons could be programmed with feedback mechanisms to manage employees such as telling them when to lift an item, where to move an item, or how fast to complete work tasks. We can conceptualize this pedagogical element as providing an increase in the active exoskeleton's relative authority, because the suit is interpreting data and providing guidance for wearers. This guidance may be used to optimize efficiency or could be used for other goals such as helping wearers increase workplace safety, help wearers control their movements, or help wearers have a more comfortable experience in the suit. We revise P3 accordingly:

> **P3:** Increases in relative authority for humans or machines will produce an increase in relative power.

P3 is also relevant for human actors in an exoskeleton adoption context. Early adopters of exoskeleton technology are likely to be seen as having cutting-edge technical expertise with these systems which is likely to increase their relative power to influence coworker behavior. This form of expertise can increase the relative authority an employee has in the workplace as other organizational members are likely to turn to these early adopters for information about the technology. In a network analysis of structural changes after technology adoption, Burkhardt and Brass (1990) found that early adopters of the technology increased their network centrality and power in their organization. In the following section we move beyond the pre-interactional phase and explore DPT propositions within interactions.

Interactional Propositions

Another major component of DPT involves predicting relationships between the perceived amount of power an individual feels they have and whether they are likely to attempt to control another's behavior in a social interaction. Dunbar (2004) argued that while dominant behaviors could constitute a control attempt, there are a multitude of other strategies that individuals can enact to control another's behavior. Multiple studies using DPT

have revealed that perceived power and control attempts show a curvilinear relationship between power and dominance, as illustrated through dominant gestures, more interruptions, and more argumentative language (Dunbar, 2004; Dunbar & Abra, 2010; Dunbar & Burgoon, 2005; Dunbar et al., 2016). In the fourth proposition of DPT (P4), Dunbar et al. (2016) argued that individuals who feel they have high or low power in an interaction are less likely to engage in control attempts when compared to individuals who perceive equal, small, or moderate power differentials in a dyadic interaction. In other words, when trying to establish the "pecking order" in an organization, a built-in hierarchy means that coworkers at the same level will vie for position through dominance behaviors with one another (such as arguing or contradicting) more than they will with people above or below them in the hierarchy. Across several studies, Dunbar has found that indeed, equal-power partners use the most dominance followed closely by the high-power partners, while low-power partners use the least dominance (Dunbar et al., 2016).

While the fourth proposition does not need revision for the exoskeleton context, we emphasize that this proposition is scalable to larger units of analysis. In work environments where exoskeletons are seen as a resource, access to the exoskeleton suit will give wearers higher levels of perceived power in work interactions with non-wearers. When non-wearers perceive wearers as having more power, this perception may shape interactions between team members or dynamics between separate teams in an organization. Consistent with the curvilinear relationship between perceived power and control attempts, we argue that suit wearers will engage in more control attempts with one another since the power distance between wearers is relatively equal when compared to the power distance between a wearer and a non-wearer. These control attempts may also be observed between teams or departments made up of exoskeleton wearers. Since wearers (as individuals or teams) are likely to perceive relatively equal levels of power between each other, they will be more likely to discipline one another, influence how one another uses the suit, or spark conflict over disagreements regarding suit use. Given the larger power gap between non-wearers and wearers, non-wearers (at the individual or team level) may be less likely to engage in control attempts over a wearer or influence how the wearer uses the suit.

The fifth through the seventh DPT propositions (P5–P7) address how actors respond to control attempts in relationships (Dunbar et al., 2016). When power differentials are high in dyads, control attempts are more likely to be successful (P5). In the exoskeleton context, wearers or the suit itself may have more power depending on organizational context. In the example of an active exoskeleton being programmed with management or surveillance capabilities, it is likely that when the suit attempts to control the wearer that the wearer will comply. Similar to P2 and P3, we argue that this proposition can be extended to human or machine actors:

> **P5:** In human-machine relationships with high power discrepancies, control attempts by the more powerful actor are likely to succeed.

In the sixth (P6) and seventh (P7) proposition, Dunbar et al. (2016) argued that in relationships where actors perceive equal power, small power differentials, or moderate power differentials, they are more likely to engage in counter-control attempts. These counter-control attempts can mitigate the control attempts that may be used to shape outcomes. We argue

that in the exoskeleton context, P6 and P7 do not need revision but can be scalable similar to P4. Exoskeleton wearers (either at the individual or team level) who engage in control attempts with other exoskeleton wearers are likely to be met with counter-control attempts. Counter-control attempts mitigate the risk of one wearer or team from dominating their peers and may be needed to keep power balances in check.

Post-Interaction Propositions

The post-interactional phase in DPT addresses long-term impacts of power dynamics or power imbalances in relationships (Dunbar et al., 2016). Dunbar (2004) argued that relational satisfaction is a key variable that can be explained when power imbalances are investigated in long-term relationships. In the eighth proposition (P8) Dunbar et al. (2016) argued that perceived power has a curvilinear effect on satisfaction in relationships. This means that actors who perceive their power as extremely low or high will be less satisfied in their relationships when compared to relationships where power differentials are small or moderate. Although relationship satisfaction is more applicable to interpersonal dyadic relationships, we argue that power imbalances can impact key variables on the team or organization level such as team satisfaction or job satisfaction. In an organizational context, perceived power might not have the same curvilinear effect on job satisfaction as it has in relational satisfaction, because there are already large power discrepancies between organizational members (e.g., CEOs when compared to subordinate employees) in the workplace. However, perceived power imbalances, such as loss of autonomy, can have negative impacts on workers by increasing stress levels and reducing job satisfaction. For instance, Mahon (2014) found that nurses in lower positions of power felt less respected by fellow hospital employees, which led to higher feelings of stress, and ultimately had higher levels of attrition than nurses with more power. As these workers gained more knowledge, experience, and autonomy, they reported feeling less of a power imbalance and a higher intention to stay in their current position. Being under constant surveillance in an exoskeleton can make employees feel that they have less power and control than the technology that is being used to manage their performance. This power imbalance may lead workers to employ resistance tactics to find balance. Introducing new technologies into the workplace that fundamentally change the work being performed or make a worker's current skills obsolete will undoubtedly cause shifts in power, added stress, and uncertainty for workers as well. We revise P8 in the following way:

> **P8:** Perceived power imbalances between humans and machines will have a negative effect on overall job satisfaction.

In the following section we explore the concept of power domains in an exoskeleton context; power domains allow researchers to explore additional outcomes of power balances in relationships including conflict.

Power Domain Propositions

In long-term close relationships, DPT treats power as generalizable across a relationship. For example, a team member with expertise in accounting might have more power when

making decisions about financial resources whereas an employee with expertise in design might have more power over product development, but both employees will average out to be a relatively power-balanced relationship. Dunbar et al. (2016) argued that, "the degree to which the domain contributes to that power dynamic depends on the importance of that domain to the relationship" (p. 86). In their ninth proposition (P9), Dunbar et al. argue that when particular domains are considered valuable, the dyad can be conceptualized as having higher power balances. Dunbar et al. also argued that a higher power balance is likely to make the actors more interdependent with one another. Since the amount of control attempts are correlated with the length of a relationship between actors, DPT predicts that a high-high power balance in the dyad will be positively associated with control attempts (Dunbar et al., 2016). We argue that P9 applies in this context as well because in organizations where exoskeleton technology is considered valuable, workers who use exoskeletons will have higher power balances and are likely to engage in more control attempts with one another.

In DPT, power domains can impact whether conflict is likely to arise in dyadic interactions. In their tenth proposition (P10), Dunbar et al. (2016) argued that in relationships with balanced power that conflict is more likely to occur in domains in which both actors have power rather than in domains in which only one actor has high power. It is likely that team members of equal status who have high power in the exoskeleton domain are more likely to engage in conflict over use of this technology than team member interactions in which only one member has power in the exoskeleton domain. We argue that P10 does not need revision and will hold true in the exoskeleton context. Similarly, P11 and P12, regarding culture and gender according to Dunbar et al. (2016) are not revised for this context.

A recent revision of DPT (Dunbar et al., 2016) argued that people negotiate the domains in which they want more power. If one team has more expertise in a domain that a group values (such as technical expertise with an active exoskeleton) then that should translate into more power in the relationship generally. Even a team member who has lower status (i.e., employees in nonsupervisory positions) than another might be able to exert more power in certain circumstances in which their expertise is valued. In a study on the fluidity of power dynamics on cross-functional teams, Aime et al. (2014) found that, "the expression of power actively shifts among team members to align team member capabilities with dynamic situational demand can enhance team creativity" (p. 327). While Aime et al. were not using the DPT concept of generalized power, their findings described a similar process in which team members had more power when work situations demanded their expertise or abilities. The finding that teams that embrace situational shifts in perceptions of power, expressions of power, and the legitimacy of power expressions were able to be more creative suggests that there may be other positive benefits for teams that embrace power based on situational needs or expertise instead of treating power as a fixed attribute based on position. Given this logic, we propose an additional proposition to DPT (P14):

> **P14:** Teams with the agility to shift power domains between members will increase their effectiveness.

Janss et al. (2012) provided another example of generalizable power domains from the world of medical action teams. These teams often are formed on an ad hoc basis with

various multidisciplinary team members (physicians, surgeons, anesthesiologists, nurses) performing pre-determined roles, but leadership is often dynamic based on the situation. When there is a conflict within the team about how to proceed, it can hamper the team's effectiveness. Team members may have a shared history that affects their perception of team power relationships and they might have expertise that surpasses their official role in the team, such as an experienced nurse working with an inexperienced resident. In teams, a lower status team member may be more influential if they have valuable expertise. In using exoskeletons in the workplace, like any new technology, experienced team members will see their power within the group grow compared to inexperienced team members, regardless of their actual hierarchical status within the group.

Discussion

This paper represents the first attempt to extend DPT from human interactions to human-machine interactions. While DPT has centered on human relationships, there is nothing intrinsic to the theory that would limit its application to relationships between humans and nonhumans (including humans and forms of technology such as exoskeletons) and use in various organizational contexts (including multiple units of analysis within and between organizations). The shift to human-machine from human-human communication provides the opportunities for scholars to test the boundaries of human interaction theories, and to explore new dimensions of humanity as we create increasingly anthropomorphic machines.

It is important for researchers to recognize the organizational context when applying DPT to human-machine interactions in the workplace. These emerging technologies are not only complex in how the technology operates but also in how these technologies make people *feel*. For instance, active exoskeletons may make organizational members feel empowered or disenfranchised depending on individual perceptions formed from dyadic interactions with the suit and coworkers. Additionally, there may be second-order implications of exoskeleton use in the workplace that impact perceptions of power balances in the workplace. Some of these second-order implications involve to what extent active exoskeletons are used to surveil employees and the privacy and trust issues that will rise under these conditions. In this case the organizational context of surveillance practices will most certainly impact how powerful or powerless wearers feel while using the technology. Researchers who are interested in issues of power and trust in this context should pay close attention to how surveillance impacts the use of active exoskeleton technologies.

When a new technology is introduced into the workplace, a shift in power dynamics can occur as users adopt and adapt to technologies at different paces thus creating gaps in expertise (Barley, 1986; Burkhardt & Brass, 1990). Those who learn new technologies sooner can experience an increase in their relative authority in the organization as less knowledgeable users, including users with more legitimized authority, seek assistance or defer to the more experienced user. Early adopters of exoskeleton technologies in an industrial context will likely have a major influence on operations, and safety and training. For example, an exoskeleton user may be called upon to assist in warehouse layout changes to accommodate suit wearers, priority lists for tasks to be done with the exoskeleton, drafting safety and usage protocols, and assisting with training and adoption as early users will

be seen as "testers" for a novel technology. Accomplishing these will require an increased access to organizational resources and an elevation in relative and/or legitimate authority.

The impact on invisible power also has implications for practitioners and researchers. The exoskeleton suit is designed to augment human strength which means that people who previously were unable to work in labor-intensive environments will, with the exoskeleton, have that option. Managers implementing exoskeletons will have a much wider pool of potential candidates, and possibly reduce employment costs from workplace injuries and turnover. How will the workplace dynamics of these organizations change with more diversity, particularly age and physical sex diversity? Researchers will have the opportunity to explore what can happen when physical abilities are no longer a limiting factor in labor-intensive employment.

Finally, exoskeletons provide a novel context for studying power dynamics between a human and a machine counterpart due to the high level of interdependence. Scholars in engineering have long pondered the question of when a human should be in control versus the computer (e.g., airplane autopilot versus human captain; Draper et al., 1964 as cited in Kirkwood et al., 2021; Major & Shah, 2020). There are different answers to the control question depending on context and user preference, but what about when the technology is embodied? How does a user's opinion of the agency of the technology impact their understanding of their own power relative to the machine? Users who are suddenly capable of lifting superhuman loads will likely experience some shifts in their self-concept. When and why do some users defer to the machine to control their movements while other users insist on retaining full control? Researchers will be able to gain a more nuanced understanding of relationships between humans and machines when the machines become inseparable from the humans.

We encourage researchers to test, challenge, or extend the propositions we have proposed in order to advance knowledge of power dynamics in human-machine interactions. Research on exoskeleton adoption and human-machine interactions is in its infancy and much empirical research is needed to understand the impact of these technologies as well as the viability of DPT in human-machine research. This paper is an early attempt at helping guide research in this area. We hope that the ideas and provocations within it are helpful for researchers interested in wearable technology, human-machine interactions, and the intersection between technology use and power dynamics in the workplace.

Author Biographies

Gavin Kirkwood (PhD, University of California Santa Barbara) is based out of the Washington Metro Area (WMA) and researches the impact of emerging technologies for a variety of stakeholders. Gavin has special expertise in artificial intelligence (AI), algorithmic management in autonomous systems, and exoskeletons designed for industrial and manufacturing industries.

https://orcid.org/0000-0002-5569-9782

J. Nan Wilkenfeld is a PhD student at the University of California Santa Barbara. She researches emerging intelligent technologies used by individuals, teams, and organizations. Her work looks at factors that influence the increasingly interdependent relationships between humans and machines, including interfaces, agency, and identity, with an eye on how these technologies might shape human communication, organizing, and broader society.

ⓘ https://orcid.org/0000-0002-1591-1910

Norah E. Dunbar (PhD, University of Arizona) is a Professor of Communication at the University of California Santa Barbara and a Fellow of the International Communication Association. She teaches courses in nonverbal and interpersonal communication, communication theory, and deception detection. She has received over $13 million in research funding from agencies such as the Intelligence Advanced Research Projects Activity, the National Science Foundation, and the Department of Defense. She has published over 100 journal articles, book chapters, and encyclopedia articles and has presented over 120 papers at national and international conferences. She is the immediate past Chair of the Communication Department at UCSB.

ⓘ https://orcid.org/0000-0002-3912-3169

Funding Acknowledgment

References

Aime, F., Humphrey, S., Derue, D. S., & Paul, J. B. (2014). The riddle of the heterarchy: Power transitions in cross-functional teams. *Academy of Management Journal, 57*, 327–352. https://doi.org/10.5465/amj.2011.0756

Banks, J., & de Graaf, M. (2020). Toward an agent-agnostic transmission model: Synthesizing anthropocentric and technocentric paradigms in communication. *Human-Machine Communication*, *1*, 19–36. https://doi.org/10.30658/hmc.1.2

Barley, S. R. (1986). Technology as an occasion for structuring: Observations on CT scanners and the social order of radiology departments. *Administrative Science Quarterly, 35*, 61–103.

Beane, M. (2019). Shadow learning: Building robotic surgical skill when approved means fail. *Administrative Science Quarterly, 64*, 87–123. https://doi.org/10.1177/0001839217751692

Boudreau, M. C., & Robey, D. (2005). Enacting integrated information technology: A human agency perspective. *Organization Science, 16,* 3–18. https://doi.org/10.1287/orsc.1040.0103

Burkhardt, M. E., & Brass, D. J. (1990). Changing patterns of patterns of change: The effects of a change in technology on social network structure and power. *Administrative Science Quarterly, 35,* 104–127. https://doi.org/10.2307/2393552

Chan, N. K., & Humphreys, L. (2018). Mediatization of social space and the case of Uber drivers. *Media and Communication*, 6, 29–38. http://doi.org/10.17645/mac.v6i2.1316

de Looze, M. P., Bosch, T., Krause, F., Stadler, K. S., & O'Sullivan, L. (2016). Exoskeletons for industrial application and their potential effects on physical work load. *Ergonomics, 59,* 671–681. https://doi.org/10.1080/00140139.2015.1081988

Demaitre, E. (2019, August 8). The robot report August 2019 issue on exoskeletons. *The Robot Report*. https://web.archive.org/web/20191019072153/https://www.therobotreport.com/exoskeletons-robot-report-august-2019-issue

Dunbar, N. E. (2004). Dyadic power theory: Constructing a communication-based theory of relational power. *Journal of Family Communication, 4*, 235–248. https://doi.org/10.1080/15267431.2004.9670133

Dunbar, N. E. (2015). A review of theoretical approaches to interpersonal power. *The Review of Communication, 15*, 1–18. https://doi.org/10.1080/15358593.2015.1016310

Dunbar, N. E. (2016). Dominance and submission in family dynamics. In J. Lebow, A. Chambers, & D. Breunlin (Eds.), *Encyclopedia of couple and family therapy*. Springer.

Dunbar, N. E., & Abra, G. (2010). Observations of dyadic power in interpersonal interaction. *Communication Monographs, 77,* 657–684. https://doi.org/10.1080/03637751.2010.520018

Dunbar, N. E., & Burgoon, J. K. (2005). Perceptions of power and interactional dominance in interpersonal relationships. *Journal of Social and Personal Relationships, 22,* 207–233. https://doi.org/10.1177/0265407505050944

Dunbar, N. E., Lane, B. L., & Abra, G. (2016). Power in close relationships: A dyadic power theory perspective. In J. A. Samp (Ed.), *Communicating interpersonal conflict in close relationships: Contexts, challenges and opportunities* (pp. 75–93). Routledge. https://doi.org/10.4324/9781315774237

Ekso Bionics. (2019). Endurance you can wear. https://web.archive.org/web/20190704033439/https://eksobionics.com/eksoworks/eksovest/

Fortunati, L. (2014). Media between power and empowerment: Can we resolve this dilemma? *The Information Society*, 30(3), 169–183. https://doi.org/10.1080/01972243.2014.896676

Fortunati, L., & Edwards, A. P. (2020). Opening space for theoretical, methodological, and empirical issues in human-machine communication. *Human-Machine Communication*, 1, 1–18. https://doi.org/10.30658/hmc.1.1

Fujiwara, K., Hoegen, R., Gratch, J., & Dunbar, N. E. (2021). Synchrony facilitates altruistic decision making for non-human avatars. *Computers in Human Behavior*, 128, 107079. https://doi.org/10.1016/j.chb.2021.107079

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

Giacomin, J. (2014). What is human centered design? *The Design Journal, 17*(4), 606–623. https://doi.org/10.2752/175630614X14056185480186

Gibbs, J., Kirkwood, G., Fang, C., & Wilkenfeld, N., (2021). Putting the AI in team: Reconceptualizing team processes when technologies are teammates. *Human Machine Communication, 2,* 153–171. https://doi.org/10.30658/hmc.2.8

Guzley, R. M., Dunbar, N. E., & Hamel, S. A. (2002). Physician-patient communication and technology: 21st century interface. *Communication Yearbook, 26,* 326–364. https://doi.org/10.4324/9781410606662

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.

Harrison, D. A., & Klein, K. J. (2007). What's the difference? Diversity constructs as separation, variety, or disparity in organizations. *Academy of Management Review*, *32*, 1199–1228. https://doi.org/10.5465/amr.2007.26586096

Horaczek, S. (2020, February 28). Robotic exoskeletons are storming out of sci-fi and onto your squishy human body. *Popular Science.* https://web.archive.org/web/20200229034405/https://www.popsci.com/story/technology/sarcos-exoskeleton-delta-guardian-xo/

Huber, G. P (1990). A theory of the effects of advanced information technologies on organizational design, intelligence, and decision making. *Academy Management Review, 15*, 47–71. https://doi.org/10.5465/amr.1990.4308227

Islam, M. R. U., & Bai, S. (2020). Effective multi-mode grasping assistance control of a soft hand exoskeleton using force myography. *Frontiers in Robotics and AI*, *7*. https://doi.org/10.3389/frobt.2020.567491

Janss, R., Rispens, S., Segers, M., & Jehn, K. A. (2012). What is happening under the surface? Power, conflict and the performance of medical teams. *Medical Education*, *46*, 838–849. https://doi.org/10.1111/j.1365-2923.2012.04322.x

Jehn, K. A., & Bezrukova, K. (2004). A field study of group diversity, workgroup context, and performance. *Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior*, *25*, 703–729. https://doi.org/10.1002/job.257

Jones, M. (1999). Information systems and the double mangle: Steering a course between the Scylla of embedded structure and the Charybdis of strong symmetry. In T. J. Larsen, L. Levine, & J. I. DeGross (Eds.), *Information systems: Current issues and future changes*. IFIP (pp. 287–302).

Kara, D. (2018, Dec 6). Industrial exoskeletons: New systems, improved technologies, increasing adoption. *The Robot Report.* https://www.therobotreport.com/industrial-exoskeletons/

Kirkwood, G., Otmar, C., & Hansia, M. (2021). Who's leading this dance? Theorizing automatic and strategic synchrony in human-exoskeleton interactions. *Frontiers in Psychology, 11.* https://doi.org/10.3389/fpsyg.2021.624108

Knight, J. F., & Baber, C. (2005). A tool to assess the comfort of wearable computers. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 47,* 77–91. https://doi.org/10.1518/0018720053653875

Komter, A. (1989). Hidden power in marriage. *Gender & Society, 3*, 187–216. https://doi.org/10.1177/089124389003002003

Kong, X. T. R., Luo, H., Huang, G. Q., & Yang, X. (2019). Industrial wearable system: The human-centric empowering technology in Industry 4.0. *Journal of Intelligent Manufacturing*, *30*(8), 2853–2869. https://doi.org/10.1007/s10845-018-1416-9

Kopp, C. (2011, September). Exoskeletons for warriors of the future. *Defence Focus.* http://mail.ausairpower.net/PDF-A/DT-Exoskeletons-Sep-2011.pdf

Lupton, D. (2020). Wearable devices: Sociotechnical imaginaries and agential capacities. In I. Pedersen & A. Illiadis (Eds.), *Embodied technology: Wearables, implantables, embeddables, ingestibles.* MIT Press.

Mahon, P. R. (2014). A critical ethnographic look at paediatric intensive care nurses and the determinants of nurses' job satisfaction. *Intensive and Critical Care Nursing, 30*, 45–53. https://doi.org/10.1016/j.iccn.2013.08.002

Major, L., & Shah, J. (2020). *What to expect when you're expecting robots: The future of human-robot collaboration.* Basic Books.

McGowan, B. (2018, October 1). Industrial exoskeletons: What you're not hearing. *Occupational Health & Safety.* https://web.archive.org/web/20181010213612/https://ohsonline.com/articles/2018/10/01/industrial-exoskeletons-what-youre-not-hearing.aspx

Murphy, A. (1998). Hidden transcripts of flight attendant resistance. *Management Communication Quarterly, 11,* 499–535. https://doi.org/10.1177/0893318998114001

Nowak, K. L., & Biocca, F. (2003). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators & Virtual Environments*, *12*(5), 481–494. https://doi.org/10.1162/105474603322761289

Orlikowski, W. J. (2000). Using technology and constituting structures: A practice lens for studying technology in organizations. *Organization Science, 11*, 404–428. https://doi.org/10.1287/orsc.11.4.404.14600

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.

Rosenblat, A. (2019). *Uberland: How algorithms are rewriting the rules of work.* University of California Press. https://doi.org/10.3917/res.216.0249

Sarcos Corp. (2019). Guardian XO full-body powered exoskeleton. https://web.archive.org/web/20191212170839/https://www.sarcos.com/products/guardian-xo-powered-exoskeleton

Upasani, S., Franco, R., Niewolny, K., & Srinivasan, D. (2019). The potential for exoskeletons to improve health and safety in agriculture: Perspectives from service providers. *IISE Transactions on Occupational Ergonomics and Human Factors*, *0,* 1–8. https://doi.org/10.1080/24725838.2019.1575930

Zaroug, A., Proud, J. K., Lai, D. T. H., Mudie, K., Billing, D., Begg, R. (2019) Overview of computational intelligence (CI) techniques for powered exoskeletons. In B. Mishra, S. Dehuri, B. Panigrahi., A. Nayak., B. Mishra, & H. Das (Eds.), *Computational intelligence in sensor networks* (pp. 353–383). Springer. https://doi.org/10.1007/978-3-662-57277-1_15

Zhang, J., Fiers, P., Witte, K. A., Jackson, R. W., Poggensee, K. L., Atkeson, C. G., & Collins, S. H. (2017). Human-in-the-loop optimization of exoskeleton assistance during walking. *Science*, *356*(6344), 1280–1284. https://doi.org/10.1126/science.aal5054