UNIVERSITY OF CALIFORNIA
SANTA CRUZ

HOW COLLABORATIVE SYNCHRONY AFFECTS HUMANIZATION OF MACHINES AND REHUMANIZATION OF HUMANS

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

PSYCHOLOGY

by

Alina S. Larson

June 2019

The Dissertation of Alina S. Larson is approved:

____
Professor Jean E. Fox Tree, chair

____
Professor Alan Kawamoto

____
Professor Leila Takayama

____
Lori G. Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures

Abstract

How Collaborative Synchrony Affects

Humanization of Machines and Rehumanization of Humans

Alina S. Larson

Whether another person is perceived as similar or different from oneself can lead to differing impressions of their inherent human-like qualities (Haslam, 2006). As the boundaries between human-computer interaction and computer-mediated communication become increasingly fuzzy, it is important to assess in which ways these humanization measurements may also apply to non-human agents. There is evidence for interventions involving motion synchrony that can strengthen interpersonal bonds and perhaps instill a sense of similarity between individuals who otherwise have little in common (Wiltermuth & Heath, 2009), but it is as yet unclear whether similar effects can be observed for interactions between human and non-human agents. In a series of three experiments, I investigate the effect of synchrony on interactions with machine agents in three health contexts (mental, physical, and social health), each designed to be successively more applicable to real-life interactions between humans and machines (using static robot therapists, a physical robot exercise coach, and then an anonymous on-line debate and a game-like activity with a supposedly human or machine addressee). Humanization ratings did correspond to robot preference, and participation in an optional mimicry task after the initial training did lead to more positive humanization scores for the robot that was designed to help, rather than replace human therapists (Experiment 1). Synchronized stretching exercises resulted in warmer responses during interactions with a physical robot than non-synchronous exercises (Experiment 2). People gave higher humanization ratings to supposedly human agents, and are more persuaded by synchrony with anonymous human agents or non-synchrony with machine agents (Experiment 3). This research provides evidence for the importance of synchrony and backstories in building positive interactions between humans and machine agents, as well as between humans and other humans in anonymized contexts. However, it is important to bear in mind that attitudes around the capabilities of machines as compared to humans can lead to different responses following synchrony exercises. For example, while participants are more persuaded by emotionally distant machine agents, non-threatening outgroup members receive greater humanization benefits from motion-matching activities.

Acknowledgements

First of all, I would like to thank my advisor, Jean E. Fox Tree. I feel incredibly lucky to have had the chance to work under the guidance of such an excellent mentor, who has offered unwavering support throughout my studies. I would also like to thank those who served on my qualifying and/or dissertation committee: Alan Kawamoto, Leila Takayama, Grant McGuire, and Travis Seymour (who also was a great help with Python advice for Experiment 1, and provided constructive feedback on an earlier version of this dissertation work). Special thanks to Robert P. Galloway for his work in writing Python scripts to help me restructure data sets for Experiment 2, as well as for the GUI used in Experiment 3. His support has meant the world to me.

Additionally, I would like to thank my research assistants who helped me with these three experiments: Daisy Iglesias, Madison Zaletel, Roselynn Hall, Madison Freitas, Griffen Bartholomew, Madison Colby, Mandeep Matharu, Tyler Bartholome, and Lydia Arce. I would also like to thank the members of my lab for their feedback and friendship throughout my time at UC Santa Cruz, and for the support of my cohort of graduate students: Jennifer Day, Peter Krause, Julia Soares, and Chris Karzmark.

Finally, I would like to extend a big thank-you to my friends and family, as well as members of the celtic music and social dance communities around the Bay Area for their part in encouraging a healthy work-life balance while in this graduate program.

How Collaborative Synchrony Affects Humanization of Machines and Rehumanization of Humans

The prevalence of machines with increasingly humanlike interaction styles has given rise to questions and concerns over competition with robots for jobs and the threat of automation in the workforce. There is at least one social parallel in the fear of having to compete with machines or immigrants for jobs. To find ways to improve interpersonal relationships and promote intergroup empathy, it is helpful to study interactions with non-human agents as a proxy for human-human interactions. The study of human-robot interactions may tell us something that human-human interaction studies cannot. For example, participants might be more willing to discuss prejudices against machines than against people in survey responses. I have developed a program of research in order to better understand which situational factors (internal and external) promote the perception of illusory personhood and self-identification with machines, with the hope that these findings may be extended to prejudice reduction in human-human interactions as well.

Previous research suggests that promoting a sense of similarity between individuals can result in more positive interactions and intergroup empathy, whether that similarity is based on internal or external characteristics. The more similar to the user, helpful, and humanlike an agent is described as being in a given backstory, the more the user enjoys the interaction and respects their feedback, even when the user knows they are interacting with a lifeless computer (Nass & Brave, 2005). The more people act in synchrony with one another in external activities, through rhythmic (even joyless) behavior-matching and collaborative motion, the more they identify as team members and the better they perform in cooperative tasks (Wiltermuth & Heath, 2009). I aim to discover whether impressions of an agent's internal characteristics and cooperation with them can be altered by manipulating the external behavior of a machine agent (using motion synchrony), even when their backstory presents their internal characteristics in a more negative light (as less similar to the identity of the participant).

In this introduction, I will first discuss how people form first impressions of personality through multimodal cues, and how they use these biased impressions to make dehumanizing judgments. Next, I will explore human-computer interactions and how it is that people come to see machines as social agents and even as humanlike, and what factors contribute to increased humanization of inanimate machine agents. Finally, I will examine the many factors that play into human-human communication and getting thoughts across through auditory and visual language – how people speak to benefit their listener – and will discuss some synchrony exercises that may improve intergroup relations.

Impression Formation

Much of human communication goes well beyond language. Even before we come in contact with someone in face-to-face conversation, we can already start to form impressions of their character. People are capable of making snap judgments to determine whether someone is human or not, a friend or not. These dispositional attributions may contribute to predictions that prove very useful for survival, but can also serve to reinforce a sense of *self* versus *other* and *us* versus *them* that in turn produces stereotype, stigma, and intergroup tension.

However, while impression formation may be adaptive in certain circumstances (life or death), this comes at some cost to accuracy. Working to better understand how dispositional thinking takes root in first impressions may allow us to be better aware of our own biases in order to have clearer communication with others and mutually beneficial interactions. In this review, I take into consideration the various multimodal cues that factor into person perception and intergroup stigma as well as ways to counter these biases and prejudices.

In the following section, I will first discuss how person perception is prone to inherent biases that are context-dependent and formed very quickly. Next, I will describe two major forms of dehumanization that can result from these character impressions – animalistic and mechanistic dehumanization – and how these two types of dehumanization are used to characterize and stereotype liberals and conservatives respectively. Finally, I will discuss potential methods to disrupt the process of negative impression formation and to increase intergroup empathy.

Bias-Prone Impressions

Knowing who may pose a potential threat to one's own survival is a highly adaptive skill, and it is clear that humans excel at doing just this, even with very limited visual cues to work from. In a study designed to see just how fast a person could judge a neutral expression for certain qualities, with accuracy measured by a comparison to judgements made by other participants after a longer exposure, researchers found that participants could decide in 39 milliseconds whether a face was threatening (Bar, Neta, & Linz, 2006). However, participants were not able to judge the same faces for intelligence in this amount of time. The authors conclude that perhaps having an understanding of threat takes priority over certain other qualities. That, or intelligence may be more difficult to identify, even from stereotypes.

These snap judgments also extend to the world of politics, as demonstrated in a study designed to determine how the appearance of political candidates factored into their electoral success (Olivola & Todorov, 2010). The researchers found that perceived facial competence was by far the strongest predictor of election success (for both hypothetical and actual votes), even controlling for other facial features that were highly correlated with competence itself. The authors were also able to construct a computational model of face-based competence judgments in order to derive specific facial features associated with competence, which is clearly rooted in subtle, visual cues (such as facial angularity, brow height). The authors found that such subtle cues can lead participants to cast a *vote* after only 100 ms of viewing a facial stimulus for a given politician (regardless of that candidate's actual competence score). Enough of these snap judgments could indeed determine the results of an election – a sobering thought.

It is clear that visual cues such as apparent facial competence may play a role in the election of our leaders, and there is also evidence that visual cues play a role in the selection of subordinates. In a mock interview study, a participant's judgments and hiring decisions were shaped by whether a female interviewee demonstrated a dynamic, authentic smile, as opposed to a dynamic and fake smile, or a neutral expression (Krumhuber, Manstead, Cosker, Marshall, & Rosin, 2008). These dynamic and authentic smiles lead to more favorable job, person, and expression ratings for both human and synthetic expressions. Female interviewees performing these expressions were also more likely to be short-listed and selected for the job in both human and synthetic face experiments. These results make a clear case that these visual cues are so fundamental to our impression formation of others that it is worth investigating whether they extend to humanoid machine agents as well.

As for other visual features that can be assessed quickly, some have found that race (Black/White) can be assessed in as little as 100 ms, and gender (Male/Female) can be assessed after only 150 ms of stimulus onset (Ito & Urland, 2003). Acoustic features may also factor into attributional inferences. For example, high and intermediate subjective measurements of loudness and resonance of voice are associated with higher vocal attractiveness for male and female voices, although objective spectrogram measures that could predict vocal attractiveness for male voices failed to do so for female voices (Zuckerman & Miyake, 1993).

It is clear that subtle multimodal cues can play a large role in the formation of first impressions. These features can be assessed from recordings, photographs, and can be made in milliseconds. It is likely that group categorization and quick assessment of others have played some role in the survival of our predecessors, but the question is: how much of the time can these impressions that are made so quickly, off of such impoverished information, be accurate?

Something to consider when evaluating first impression accuracy is whether someone has the same impression of another person, regardless of that person's mood in the present moment. In one study, researchers examined this question by having participants come into a lab to rate five photographs each of twenty people in order to see if there was within-individual variation in impression formation. The authors demonstrated that snapshots of dynamic facial expressions can lead to wildly different impressions, with participants more likely to be consistent with other participants for a specific image than with themselves for the same face making different expression (Todorov & Porter, 2014). Additionally, the authors demonstrated that preferences for a given face were also different in different contexts (online dating as opposed to mayoral elections). This finding is further evidence that the context of an introduction and the expression of a face when one first sees it can alter dispositional attributions one makes.

Once biased impressions take shape, they can be difficult to shake. In one study of group categorization and disclosure, participants listened to an audio recording where a male speaker disclosed his sexual orientation either early or late in the recording (Buck & Plant, 2010). Male participants who heard the speaker disclose their orientation as gay earlier in the recording responded with more negative and avoidant behaviors than those who heard the disclosure later in the monologue. The recorded voice who disclosed as gay was met with more prejudice overall than that of the same voice who disclosed as straight, though there was a clear primacy effect for sexual disclosure prejudice. In other words, higher level of bias occurs when minority-group categorization occurs before individuating information. Initial impressions carried greater weight in ongoing character analysis.

It is also important to ask how the visual context surrounding a judgment might work to influence the type of attributions made. In another experiment**,** student mock jurors viewed one of three videotapes of the same mock interrogation, resulting in a confession by the suspect (Lassiter & Irvine, 1986). In the suspect-focus group (where the suspect was in full view of the camera), students could recognize a small degree of coercion used by the detective, but made more dispositional attributions of the suspect than in the other two groups. In the detective-focus group, students noticed a large degree of coercion, and made the fewest dispositional attributions of the suspect. The third angle, filmed from the side, such that suspect and detective were in equal-focus, allowed moderate recognition of coercive behaviors on the detective's part.

In order to validate the findings in this initial study, as well as to challenge aspects of it that could fall under critique, a number of replication studies were performed. In four additional experiments, the authors concluded that the equal-focus angle was the least biased camera angle, and that people are still prone to camera-angle bias even when they are aware of the impact of viewing angles, or when told to focus more on the evidence provided in a case (with likelihood of guilty verdicts and term sentences increasing as the camera angle focused from detective to suspect). This perceptual-level bias, which the authors describe as a form of *mental contamination* seems difficult to override and prevents participants from noticing coercive techniques used by the interrogator (Lassiter, Beers, Geers, Handley, Munhall, & Wiland, 2002).

Another couple of attempts were made in more ecologically valid contexts, but neither a realistic video of a trial re-enactment nor intensive judicial instruction of mock jurors were able to eliminate the influence of camera angle on a mock juror's final verdict (Lassiter, Geers, Handley, Weiland, & Munhall, 2002). The authors recommend equal-focus or detective-focus camera angles for interrogation footage, which has a tendency to allow viewers to pick up on coercion more effectively.

In summary, the literature suggests that many subtle cues related to a person's appearance, their placement within a scene, as well as the timing of information presentation can have lasting effects on attributions made about another individual. These judgments, in turn, can have life-or-death consequences for the individuals under scrutiny, as in the cases of actual courtroom verdicts made by jurors.

Dehumanization

Dispositional impressions formed in the ways described above might in turn lead to dehumanizing characterizations of a perceived individual. Dehumanization is thought to take one of two major forms (Haslam, 2006). The first form is that of *animalistic dehumanization*, denying the subject their fundamental traits of *human uniqueness*, such as reasoning capability, maturity, moral sensibility, refinement and civility. The second form is that of *mechanistic dehumanization*, denying the subject fundamental traits of *human nature*, such as emotional capability, warm and deep interaction capability, open mindedness, and agency.

A number of factors can affect whether or not dehumanization occurs. In a series of experiments, researchers compared the effect of communication medium (written as compared to spoken, and later compared to synthesized speech) and agreement (whether the participant agreed or disagreed with the communicator) on perceptions of their interlocutors (Schroeder, Kardas, & Epley, 2017). When participants agreed with their interlocutors, communication medium did not have consistent effects on humanization ratings of communicators. However, in cases of disagreement, communicators who spoke were rated much more favorably than those who communicated the same information via written text. These results clearly indicate that contexts with written as compared to spoken word and where participants disagree with their interlocutor on the topic of communication can lead to lower ratings of human uniqueness and human nature qualities. In Experiment 3 of the present research, participants are set up to dehumanize an interlocutor in order to stage an intervention. All participants read text that is selected to be at odds with their own opinion. The fact that the text is read rather than spoken, and that the interlocutor is in disagreement with them will help to ensure that participants begin to think of the addressee as an outgroup member.

In terms of humanization subscales, it is interesting to note that the authors observed reduced effects of communication medium on human nature as compared to human uniqueness ratings across all experiments (Schroeder, Kardas, & Epley, 2017). They suggest that this could be a result of spoken language leading to judgements on thinking and cognition (human uniqueness) more than emotion and feeling (human nature). In my own research, since all participants in Experiment 3 are exposed to written (rather than spoken) language that disagrees with their own argument, I will not be so concerned with these comparisons, but it is interesting to see how different cues can give rise to different attributions along the two humanization subscales. Human uniqueness and human nature characteristics can manifest differently in different contexts and can have social consequences when first impressions become lasting impressions.

These two forms of dehumanization have a demonstrated a role in partisan politics. In a Mechanical Turk experiment, participants were asked to rate liberals and conservatives for positive and negative human uniqueness and nature traits. Next, participants were asked to identify their own political ideology. Human uniqueness and nature traits were associated with conservatives and liberals respectively (Crawford, Modri, & Motyl, 2013). That is, conservatives associated with terms related to reasoning, but not emotionality (mechanistically dehumanized) while liberals were associated with emotionality, but not reasoning (animalistically dehumanized). Additionally, more positive descriptors were applied to the participant's own political party (in-group), while more negative descriptors were applied to the opposing party (out-group).

In another study of political identity and dehumanization, researchers found that the more strongly participants identified with a political ingroup, the more they would dehumanize an outgroup (Pacilli, Roccato, Pagliaro, & Russo, 2015). Not only are people likely to apply negative animalistic or mechanistic labels to an outgroup, but they will also apply positive human uniqueness and nature traits to their own group, and the more strongly they associate with this ingroup, the more they will consider the outgroup to be less than human. While it may seem hopeless to expect to counter these snap judgments that people make, there are some areas where potential interventions might prove effective that I will discuss in the next section.

Potential Interventions

While uncommon, it seems that people are able to impose critical thought that is independent of first impressions when they are asked to do so. In a mock trial study, arguments for and against the defendant's claim to innocence were presented to a participant (Tetlock, 1983). The presentation order for pro/con arguments varied, as did the participant's requirement to justify their decision or not, and whether they were asked to justify their decision before or after reading the arguments. Participants who received the arguments against the defendant first (con/pro) were more likely to perceive the defendant as guilty than those receiving arguments in favor of the defendant first (pro/con). However, order of argument presentation made no difference when participants were asked to justify their decisions before viewing the arguments.

Others have demonstrated that facial competence cues only impact election success for the less knowledgeable voters, with no impact on highly knowledgeable voters (Olivola & Todorov, 2010). The more situational knowledge a person has, and the more they are encouraged to think critically, the less influence dispositional cues have in final judgment. Becoming more knowledgeable can take the form of exposure to others different from oneself.

There is evidence to suggest that intergroup contact with others from diverse backgrounds can help bolster empathy and lead to reductions in prejudice between groups. In a discussion of intergroup contact theory, some researchers postulate that there is an inverse relationship between intergroup contact and prejudice, and that familiarity does indeed have a tendency to increase liking (Pettigrew & Tropp, 2006). Other researchers have also proposed a theoretical model to help explain individual differences in punitiveness. These researchers have demonstrated an inverse relationship between empathy and harsh punishment, as people are more punitive towards those they don't empathize and identify with (Unnever & Cullen, 2009). Taken together, these theoretical models would suggest that in order to foster empathy and goodwill between groups, we would do best to encourage intergroup contact whenever possible.

Some researchers did indeed recommend intergroup contact as a method to reduce or overcome dehumanization (Haslam & Loughnan, 2014). Another recommendation they alluded to was that of promoting a superordinate identity, and focusing on similarities shared between subgroups. The authors lamented the fact that dehumanization reduction is under-researched. This could, in part, be due to the difficulty of ethically setting participants up to exhibit dehumanizing behaviors, in order to create ecologically valid settings to study the reduction of such behaviors. Such a design could potentially exacerbate existing behaviors, or lead participants to feel uncomfortable responding honestly within an experiment.

In the context of my research, intergroup contact takes the form of a collaborative synchrony or mimicry exercise, and common identity is established by way of backstory (telling people they will be interacting with another student, as compared to a machine learning algorithm). By comparing dehumanization between machine agents, or between a so-called human versus machine, I hope to investigate possible interventions with machines as proxies for humans. For this to work, it is important that machine agents can be seen as outgroup members (subordinate or threatening to humans, leading to dehumanization). Simultaneously, it is important that people are able to attribute humanlike characteristics to these machines (leading to humanization).

Machine agents can be seen as outgroup members. In a mail-in survey, participants from Northern California were asked to what extent they believed computers were capable of performing physical and psychological tasks usually assigned to humans, and whether computers could potentially fill routine, interpretive or personal occupations. Participants responded with an anthropocentric perspective – believing that robots should not fill roles traditionally held by people – more when they had either limited experience with other cultures, a lower level of education, or both (Nass, Lombard, Henriksen, & Steuer, 1995). It is possible that negative impressions of robots go beyond their actual limitations, and have more to do with the perception of threat that comes with less exposure to diverse populations.

Machine agents can also be seen as humanlike. In another survey study, asking participants about preferred occupations for robots, people indicated a preference for collaborative, as compared to competitive robot workers (Takayama, Ju, & Nass, 2008). It is likely when robots are positioned as taking on the tasks traditionally assigned to humans, people begin to think of these robots as threatening, and even as outgroup members. Perhaps an increased exposure to robots could carry with it a greater understanding of computer capabilities and reduce the sense of threat that some feel.

In sum, the body of literature on impression formation seems to indicate that presentation context has a large impact on how people engage with and judge other people, as well as on in-group and out-group categorization. In the next section, I discuss ways in which computers, robots, and other embodied agents can be subject to similar judgments.

Illusory Agency in Human-Computer Interactions

It is clear that humans are experts at formulating character impressions of other people, if not always consistently across contexts. As machine technology improves by leaps and bounds, and it becomes easier for humans to make use of smart devices, the distinction between what is *human* and what is *machine* becomes fuzzier. Consequently, people begin to ascribe humanlike characteristics to otherwise non-conscious objects, such as a smart home device that can hold up a semi-scripted conversation with them, and follow simple instructions. This in turn can lead users to *agentify* these objects and imbue them with a certain perceived consciousness. That is to say, people may be in some ways tricked into thinking of machines as real people (effectively humanizing them) even while they are aware that they are interacting with mindless machines.

In the next section, I will start off by describing a series of experiments that document ways in which humans begin to treat machines, or computers, as social agents. I will then discuss how the effect of agentifying machines becomes even stronger when the machine is made to more closely resemble the human it is interacting with. Finally, I discuss the power of belief: how thinking of an agent as human or machine is enough to change how people choose to interact with another agent, even if they are in actuality no different from one another.

Machines as Social Agents

Even while knowingly interacting with machines, people have a tendency to treat robots as something between subject and object. Because robots straddle the line between an animate *other* and an inanimate *thing*, it is unclear as to whether they are thought of as having a *mind* or *consciousness*. Various experiments have been conducted by Nass and colleagues to study the tendency that people have of treating machines as social agents, while simultaneously denying them personhood (Nass & Brave, 2005). For example, one study made use of text-to-speech (TTS) synthesized voices, a type of computer-generated speech. Participants were asked to listen to five book reviews produced either by five different TTS voices (of different pitch, timbre), or just one TTS voice for all five reviews. When participants heard all five different voices, the speakers were rated as more socially present and the books were given higher ratings as well (Nass & Moon, 2000). These effects persisted even if the experimenters carefully informed participants of what TTS speech was, how it was produced, and informed them that none of the voices were in fact reflective of actual human speakers.

In another study, when participants received flattering (as opposed to generic) feedback from a human or synthesized speaker while playing a trivia game, flattery boosted ratings of the synthesized speaker's social attractiveness for synthetic speech but not recorded human speech (Lee, 2010). Interestingly, human voices only received received slightly higher anthropomorphism scores than synthesized speech. Simply hearing a synthesized voice is enough to amplify ratings of social attractiveness and believability of the computer. Still, one might expect a visual representation of a machine to clearly demonstrate to its audience that it is not in fact a conscious living creature, and need not be treated as such.

In an experiment involving a multimodal virtual agent, PERSONAGE, researchers sought to understand if the frequency and performance of gestures would impact a virtual agent's rated extraversion (Neff, Wang, Abbott, & Walker, 2010). Indeed, increased rate of gestures as well as a more dynamic performance of gestures were both associated with an increase in perceived extraversion. If people think of a virtual agent as nothing more than the result of algorithms, with no actual identity or *self*, dynamic gestures should carry no social importance whatsoever. It would seem that a visible robot is not enough to dispel users of the illusion of agency.

Understanding what cues robots can make use of in establishing their agency is becoming increasingly important as robots become more and more ubiquitous. In one study, researchers brought a robotic baby seal named Paro into a nursing home. Paro's activity generated a good deal of conversation both *about* the robot as an object, and *towards* the robot as an interlocutor (Taggart, Turkle, & Kidd, 2005). When residents addressed Paro itself, they often spoke to it as they would a pet, and many expressed a belief that Paro could experience emotional states. When the robot was switched off, rendering it a fully inert object, the conversations with researchers were noticeably less chatty, as residents spoke and gestured less. If Paro can be treated as an emotional entity in such a context, nursing home residents who otherwise have limited access to social interactions may benefit tremendously from the presence of such a robotic agent.

Relational agents, such as robots like Paro, are designed to fostering long-term emotional and social connections with their users. In one study, researchers created a humanoid animated agent, either relational or non-relational, in order to determine the impact of having a social and emotional connection with a virtual agent in the context of behavior change coaching for exercise (Bickmore & Picard, 2005). The researchers found that interacting with a relational virtual training coach led to better user impressions and desire to continue training with the agent, as well as more emotional responses in conversation with the agent, as compared to the non-relational agent.

There are clear advantages to using machines that encourage users to engage in conversation, or even healthful physical activity. As robots become more abundant and affordable, this may allow everyone to have their own personal training coach and companion. Another study made use of a NAO robot (a small robot with a humanlike body) and designed a motivational interview to encourage participants to talk about themselves and their behaviors without fear of judgement or interruption. Participants reported feeling more comfortable divulging sensitive information and being honest with the robot (Galvão Gomes da Silva et al., 2018). If robots can encourage more honest self-disclosure, perhaps this will allow people to be more honest with themselves, by extension.

Another place robots may be able to make a positive impact on people's lives is in therapy. While robots may never be as nice to talk to as other people (Larson & Fox Tree, *in review*), there are some cases where having a robot therapist may be better than an alternative of having no one to talk to. Machine agents that communicate via text might provide an avenue for social interaction and support for individuals who either don't feel comfortable disclosing their emotions to another person, or who cannot afford to see a human therapist. A study on the conversational agent Woebot found that, compared to an eBook resource on depression, conversations with this chat bot significantly reduced symptoms of depression in users as reported in a follow-up survey (Fitzpatrick, Darcy, & Vierhile, 2017).

Robots may work well as personal training coaches, therapists, maybe even doctors, but these interactions are more complicated than a simple human-object interaction, and may begin to take on more social elements. After completing a task on one computer, people give consistently higher ratings on the performance of this computer when they are asked to fill out an evaluation on that particular computer, as compared to when they are asked to fill out an evaluation on a second computer or on a printed survey (Nass, Moon, & Carney, 1999). While people all claimed that they would have responded similarly regardless of how the survey was administered, and yet it seems that they unconsciously treated the machine they interacted with with more politeness, almost as if it were a thinking and feeling agent.

Behaviors related to politeness can also indicate a certain degree of respect and comfort with a machine agent. In a robot study, proxemics (approach distances) were used as a measure of user comfort in interacting with a machine (Walters, Syrdal, Koay, Dautenhahn, & te Boekhorst, 2008). The robot's voice was either a male, female, or non-gendered neutral synthesized voice, or the voice of the human experimenter. Though approach distances were closer (indicating greater comfort) for the human experimenter's voice condition, the fact that participants left a good deal of space between themselves and the machine is possibly indicative of the fact that they had respect for the robot's space.

Robots may prove useful as tools for behavioral change, especially considering the social effects of chat with machine agents, attachments formed with a particular machine, and interactions with physically embodied and conversational machine agents. People have been shown to think of robots less like simple objects and more like pets, or maybe even companions. In one study on a mechanical ottoman footstool, researchers found that participants interacted with the robotic chair as they would with a pet, or an object, but often a bit of both

(Sirkin, Mok, Yang, & Ju, 2015). Additionally, participants used agentive language to describe what the ottoman might want as a way of explaining its behavior. The researchers used a Wizard of Oz design (with a human confederate acting as the machine), where an experimenter was in reality controlling the ottoman via remote control from another room. While most participants correctly suspected the chair was teleoperated, many of them still ascribed desires, volition, and agency to the chair itself. Some participants were so taken in by the illusion of life that they didn't feel comfortable placing their feet on the ottoman and treating it like an object. Because it acted like it had a life of its own, participants felt that it deserved better treatment than that.

Similarity Attraction

There is another clear trend in the human-computer interaction literature: people prefer to interact with machines that display similar, and sometimes complementary characteristics to their own personality and to humans in general. This, of course, harkens back to the ever-relevant *uncanny valley* hypothesis (Mori, MacDorman, & Kageki, 2012). That is to say, people appreciate objects and feel affinity for them more as they come to approach a human likeness for the most part, with the exception of a certain drop in affinity as an object approaches (but doesn't quite) appear human. For example, objects such as prosthetic limbs would fall into this region; they resemble a human in appearance quite closely, almost enough to be mistaken for a true human appendage at first glance, but are simultaneously not close enough to appear human. If an object is humanlike, but misses the mark, it may be considered eerie or unnatural. It is likely that the uncanny valley is different for people of different cultural and life experiences, with some finding a given robot or device creepy and others seeing it as perfectly normal, depending on how that technology is portrayed in the media and how common it is.

In a similar vein to research with agent voices, one study on virtual embodied agents found that participants respond more socially to less embodied computer agents. Participants were either told they were interacting with an avatar controlled by a human, or that they were interacting with a computer agent in a three-dimensional environment (Nowak & Biocca, 2004). Additionally, they were shown either a high-anthropomorphic or low-anthropomorphic image (that is, appearing as an artificial human face or simply a mouth and eyes), or no image. Participants responded socially, reporting a high level of social presence (connection: being able to assess partner's reactions and feeling like it was a face-to-face meeting) for both human and computer partners, more so for low-anthropomorphic than no image or high-anthropomorphic images. Furthermore, the virtual image improved the perception of telepresence (of feeling immersed in the virtual environment), but more so with the low-anthropomorphic than high-anthropomorphic avatar/agent again. Perhaps the high-anthropomorphic faces were close, but not close enough to human to elicit favorable responses, which was why a disembodied mouth outperformed the face for social connection and telepresence.

Another experiment used videos of robots (of different appearances and behavioral styles) to lend empirical support to the uncanny valley hypothesis. In this experiment, the researchers demonstrated that people do generally respond more positively to robots with more humanlike appearance and attributes (Walters, Syrdal, Dautenhahn, te Boekhorst, & Koay, 2008). That is to say, robots with facial features received the highest ratings for extraversion, emotional stability, conscientiousness, agreeableness and intelligence. There was one exception to this finding: introverted participants and those with lower emotional stability preferred the more mechanical robots to a greater extent than did other participants. Nevertheless, when machine agents lack facial features, it is clear that most people prefer the machine to look and act more human.

In this discussion of preference for self-similar machines, it is always important to remember that individual and group differences in preference do exist, and in fact are interesting demonstrations of which communities may indeed benefit the most from the use of certain technological innovations. Another study attempted to delve into group differences in preferences for human versus synthesized singing voices (Kuriki, Tamura, Igarashi, Kato, & Nakano, 2016). The two groups were participants with and without a diagnosis for Autism Spectrum Disorder. While autistic and non-autistic groups described the music they heard in similar terms, the autistic group rated human and synthesized singing voices more similarly for naturalness, animatedness, and emotion, even while noting that the human voice was more humanlike. The autistic group did not display the clear preference for the human voice that the non-autistic group did.

In another similarity-attraction study, dominant or submissive computer agents were randomly matched to dominant or submissive participants (Nass, Moon, Fogg, Reeves, & Dryer, 1995). Without knowing their own dominant or submissive personality score, participants preferred the computer personality that matched their own. Another study demonstrated that people responded positively to a change in the computer's behavior over time, when that change was in the direction of increasing similarity to their own dominant or submissive personality (Moon & Nass, 1996). This would indicate that, for at least dominant and submissive traits, users prefer machines to portray similar personality characteristics to themselves.

In another book review study – this time with voices designed to sound introverted or extroverted – introvert and extrovert participants demonstrated strong similarity attraction (Nass & Lee, 2000). Participants rated the reviewer as well as the book review, and in both cases tended to prefer those with a TTS voice matched to their own introversion/extraversion level. A matched TTS voice was rated as more attractive, credible, and informative.

However, in a more interactive study with a visual and textual virtual agent, the stronger effect was that of complementary attraction (Isbister & Nass, 2000). Introverted and extraverted participants were either matched to a virtual agent's visual cues (posture), verbal cues (text), both, or neither. Next, they completed the Desert Survival Problem, ranking the most important items for desert survival with input from an animated virtual partner. Participants exhibited a preference for consistent characters (with self-consistent audio-visual cues indicating either introversion or extraversion) with a personality complementary, not identical, to their own. Though this finding seems contradictory to previous findings (supporting similarity-attraction), it is perhaps the interactive nature of the task and the goal of the task that gave rise to a complementary preference, with the goal of the Desert Survival task yielding advantages for diverse viewpoints, while a complementary perspective on a book review may be less useful.

In another study, an embodied virtual agent was coded to use big sweeping extraverted gestures, or constricted introverted gestures during different segments of a story, either moving from introverted to extroverted, or extroverted to introverted gestural styles over time (Tolins, Liu, Neff, Walker & Fox Tree, 2016). Participants who had scored high or low on the Big 5 personality test for extraversion–introversion listened to the embodied agent tell a story with these extraverted and introverted gestures, and then were asked to repeat back the same story segment to the agent. The gestures of the participants were recorded and transcribed, then compared to the gestural style of the agent they interacted with. Highly extraverted participants were more responsive to the agent’s changing gestures, and would use big sweeping gestures when the agent switched to using smaller gestures, indicating that divergence, rather than convergence, may be at play in these interpersonal interactions. Perhaps the participants were shifting to a complementary style in order to better engage with their embodied interlocutor.

Whether it is important for a computer agent to display features similar or complementary to one's own, it is clear that most people prefer said agent to be more humanlike than robotic, at least up to a point. Whether this preference is based on a logical desire for more common ground and mutual understanding with an interlocutor, or on sheer prejudice against outgroup members or those dissimilar to oneself is unclear. It remains an important area of inquiry in the study of human-computer interaction, and something I seek to better understand through my own research.

Backstories

While people can begin to treat machines as social agents, there remains a clear bias to favor humans over machines. In a discourse analysis experiment, participants were either told they were having a text-based conversation with a human next-door or a computer program while working together to solve the Desert Survival Problem (Schectman & Horowitz, 2003). When they believed they were communicating with a human partner, participants acted more like they would in establishing a relationship, as compared to when speaking with a presumed machine partner. For example, they spent more time writing longer comments to supposedly human partners than machine partners, and also used more relationship statements. Even though no difference actually existed between *human* and *machine* partners, the perception of one as being human made a big difference in communicative displays.

In a similarly deceptive study, participants were told that a computer-animated character they were interacting with in a social dilemma or negotiation game was either a human avatar or a computer agent (de Melo, Gratch, & Carnevale, 2014). When participants believed they were interacting with a human rather than a computer, they were less hostile – more polite – towards their virtual partner, reported more positive impressions of them (an average of fairness, trustworthiness, cooperativeness, and likability). They were also more likely to concede to their partner in an argument than when the partner was perceived to be a computer agent. In other words, thinking that they were interacting with an actual human led participants to behave in more sociable ways. While people may treat computers as agents in many ways that seem illogical, perceived *human* partners still get better treatment.

Another study using the Desert Survival Problem manipulated displays of humor by a virtual partner, which participants were told was either a human or a computer (Morkes, Kernal, & Nass, 1998). Participants gave their so-called *human* partner higher ratings for humor, and also smiled and laughed more than when interacting with a so-called *computer* partner. Again, it would seem that believing that an interlocutor is human leads to more social displays than believing that an interlocutor is machine.

In another study, labeling an interviewer as human or computer was enough to alter the participant's (here, mock interviewee's) responses during a web-based mock job interview (Aharoni & Fridlund, 2007). Participants smiled more and filled more silences with human than machine interviewers, even while subjective reports of emotional reactions were seemingly unaffected by human-status of the interviewer. It is likely that the smiles displayed for human (as opposed to machine) interviewers was less a result of happier emotions and more a communicative tool to benefit the conversational partner, as they did not correspond to higher subjective ratings of happiness.

All in all, there are clear signs that people will respond to machines socially, and generally seem to prefer computer agents that look and act more humanlike. Still, regardless of how computer agents act, the knowledge that they are *not* human is an impediment to their ability to be seen as social actors. As computers continue to gain humanlike abilities and take on human responsibilities, it will be important to discuss how they are perceived and treated as similar to or different from humans. In order to work harmoniously alongside often humanlike machines, there will be times where people may begin to embrace machines as members of their own ingroup, as team members and colleagues or even friends.

Spontaneous Communication

Spontaneous communication in humans is a multimodal phenomenon, involving not only verbal cues, but also visual cues. Moreover, these cues and linguistic signals are not performed solely for the benefit of the speaker in transmitting a message clearly and accurately, but for the benefit of the listener. The way in which the interlocutor is perceived will have an impact on the methods employed by the speaker to establish common ground (Clark, 1996). Whether a speaker feels they have properly synced with their conversation partner will impact their impressions of this interlocutor as a conscious and thinking agent. This in turn can potentially lead to more positive interactions. Certain joint activities and communication styles may produce more synchronous and empathetic interactions, with the potential to boost a sense of ingroup membership.

In this section, I begin by describing the imperfections of auditory and visual communication, explaining how their use in conjunction with one another can make up for the inherent ambiguities of each alone (an argument for studying both visual and auditory cues together, rather than separately). In the next subsection, I summarize how various characteristics of communication – including gestures, pragmatic markers, and fillers – are used as much for the benefit of the listener as they are for the speaker themself. I discuss ways in which these cues may be used in human-computer interaction as well. Finally, I describe ways in which people begin to match the behavior of an interlocutor – through motion synchrony – and how this may be beneficial for intergroup relations.

Multimodal Ambiguity

Both visual and auditory cues can be ambiguous, which is why it is important to study how they work in conjunction with one another. In an audio-only study of sarcasm, participants could differentiate posed sarcasm from non-sarcasm using verbal cues, but could not distinguish between spontaneous sarcasm and non-sarcasm (Rockwell, 2000). Sometimes, even in human-human interactions, verbal cues are not enough to convey real, intended meaning.

Although the focus of psycholinguistics is usually on the verbal aspects of spoken communication, it is clear that visual cues also play a role. In an fMRI study, an actor spoke sentences while either facing the camera or to the side, and gesturing or not (Nagels, Kircher, Steines, & Straube, 2015). Participants felt most addressed when the speaker was facing them and using gestures, which consequently led to the greatest activity of the anterior cingulate cortex (an area associated with mentalizing) and the left fusiform gyrus (also previously found to respond to body orientations). This experiment demonstrates clearly that not only are gestures important for a listener, but body-orientation also plays a role in whether the listener feels addressed or not.

Also, a study of facial expressions found higher accuracy for posed rather than natural expression identification, which sometimes fell below chance levels (Motley & Camden, 1988). Because of cases of ambiguity like this, it is important for a speaker to use all the audio-visual cues they can muster in order to convey a message that will be interpreted with the intended sentiment and meaning. If humans are not perfect at this, it is hard to imagine that computer algorithms and artificial intelligence will be able to do much better in communicating in a humanlike way. Still, with the improvement of machine learning tools, this is no longer an impossibility, so it is important to consider which communicative signals robots should be trained on to better recognize.

Certain speech behaviors may be performed primarily for the benefit of the listener. Visual cues such as gestures as well as verbal cues such as fillers, discourse markers, and hedges all play a role in keeping the listener both engaged and on the same page, especially in light of the fact that visual and auditory cues can be misinterpreted so easily.

Cues to Benefit the Listener

One clear example of a speech cue for the benefit of the listener is taken from a study on communication between blind and sighted participants, where they engaged in four Piegetian conservation tasks, describing the quantity, length, number and mass of water that was moved into various containers (Iverson & Goldin-Meadow, 2001). Sighted participants gestured to the blind participants, which indicates that gestures may indeed be self-serving. However, all of the blind participants did in fact make use gestures while communicating with sighted listeners. One might have expected that congenitally blind speakers would see no benefit in using gestures for their own speech production, but in so doing they effectively demonstrated that gestures may serve a greater purpose than simply aiding speech production.

In another study concerning the visibility and presence of an interlocutor, participants' gestures were recorded in face-to-face dialogue, telephone dialogue, or a monologue to a tape recorder (Bavelas, Gerwing, Sutton, & Prevost, 2007). Both the visibility and presence of a conversational partner had an effect on the rate of gesturing, with face-to-face yielding the highest rates of natural gestures, telephone being the second highest (though gestures were small) and recorder yielding the lowest rates of gestures, which were tiny and strange in shape. So while participants produce some gestures even in the absence of a conversational partner, they clearly produce more when they know they will be received.

In another study, participants were asked to retell a cartoon story in dyads either face-to-face or where their partner could not see them, behind a screen

(Alibali & Heath, 2001). Participants produced higher rates of fillers (*um* and *uh*) in face-to-face than when behind a screen. Participants also produced higher rates of representational gestures (depicting semantic content) in face-to-face than when behind a screen, though beat gestures (rhythmic and non-semantic) were produced at about equal rates regardless. The beat gestures could be largely self-serving, but representational gestures are here shown to be largely produced for the benefit of the conversational partner.

In a study of filler production, participants spoke to a human voice coming from behind an opaque screen, or a computer that had a recording of the same human voice playing off of it and asking simple trivia questions (Walker, Risko, & Kingstone, 2014). When another human was present in the room, behind the screen, the participant produced more fillers. A follow-up experiment tested whether the increased use of fillers in the screen condition came from the mere presence of the human in the room. The researchers found that just having a human experimenter in the room while the recording played off the computer did increase the participants' filler production, but not as much as when the participant was interacting with the experimenter behind the screen. This finding lends further credence to the hypothesis that fillers are produced (at least in part) for the benefit of the listener.

In a cooking experiment, participants observed a scene where a human or robot helper was assisting a novice human baker in baking cupcakes using different pragmatic markers (Torrey, Fussell, & Kiesler, 2013). The use of hedges and discourse markers were carefully manipulated in order to examine whether their use by an assistant – human or machine – was useful. That is to say, whether it made the baking assistant seem more considerate, likable, and less controlling. Results indicated that hedges and discourse markers resulted in more positive ratings for both humans and robots, most of all when combined. This finding further underscores the importance of these speech cues for the listener, and why a speaker (even a machine) may benefit from the use of these communicative strategies. Interestingly, the use of discourse markers had a larger positive impact on ratings of the robot than the human helper.

To more closely examine the social impact of discourse markers and fillers in a non-task-related narrative setting, as spoken by humans as compared to synthesized voices (Larson & Fox Tree, *in review*), we tested how a speaker was rated positively or negatively for a number of traits (trustworthiness, friendliness, intelligence and nervousness). Although participants showed a clear preference for human over machine speech, we did find that the synthesized voice was rated more highly for friendliness while using speech with markers compared to without. Synthesized speech was also rated more highly (on combined trait ratings) when it was believed to have come from a human with an artificial voice box, as opposed to a machine algorithm. Since we found that actual speech patterns were less important than this framing, perhaps it is less important for machines to work on perfecting speech patterns, and more important to focus on making the agents themselves relatable. It may be that the previous findings in the cooking experiment differed in that participants were watching videos of humans engaging with these robot agents while performing a complicated task, and this contributed to a sense of the robot as being more personable.

Interpersonal Benefits of Motion Synchrony

It is clear that many spoken conversational phenomena are developed for the benefit of any listener. Of similar importance is how these behaviors are shaped and customized for the benefit of a specific listener. While engaging with a communicative partner, people may begin to exhibit signs of behavioral mimicry (convergence, synchrony, or alignment) with their partner's linguistic behavior (such as speech rate, pauses, accent, utterance length, phonology), as well as gaze and facial expressions (Branigan, Pickering, Pearson, & Mclean, 2010).

People coordinate behaviors through joint attention, perception, and action. For this to be possible, it is helpful to have shared task representations, knowing where and what others are doing, and distinguishing cause and effect of one's own actions as compared to the actions of others (Sebanz, Bekkering, & Knoblich, 2006). Moving in synchrony with another can lead individuals to feel that they are contributing to the same enacted effect, which in turn may lead to increased feelings of affinity as it becomes difficult to distinguish oneself from an other. That said, it is unclear whether this blurring of self-other distinctions relies on theory of mind, or that the other be considered a thinking and feeling, even living agent. To understand this, it is helpful to conduct synchrony studies with machines.

Synchrony, or entrainment, involves coordinated and time-locked behavioral mimicry (doing what someone else is doing, at the same time as they are doing it). Synchrony has proven so effective in helping boost team morale that it is often incorporated into games. One such game is Yamove! – a game where players wear apple phones on wristbands that measure movement intensity, synchrony, and diversity of movement (Isbister, 2012). Pairs of players get cooperative scores, and teams face each other and take turns in short rounds, with higher scores going to teams who move more in-sync with each other. This set-up is designed to feel more interactive than having everyone stare at a screen. This style of gaming may become more popular as wearable technology improves, turning player attention back towards other players and away from screens or tables. While this game is an excellent example of team-based synchrony in gaming, the focus is on the other humans in the game, not a virtual machine agent. Perhaps if a machine agent were framed as a team member that a human player could synchronize and perform with, rather than against, it might be possible to observe a blurring of self-other distinctions and a heightened sense of team membership with the machine agent.

In one series of experiments, people engaged in various exercises that have been shown to boost synchrony (Wiltermuth & Heath, 2009). Among these are joyful and joyless activities, and muscle-based and rhythm-based activities. In one study, participants walked around, either in-step with each other or not, and played a trust game. Though walking in sync didn't boost joy ratings, it did increase cooperation in the trust game. In another iteration of this study, participants who kept up a rhythm or sang a song (or both) in sync with a group showed higher cooperation and had an increased sense of team membership, though they did not report more joy.

Coordinated attention is very important for synchrony and entrainment. In one study, participants were asked to move their hand up and down in time with a metronome while an experimenter either moved their hand in the same way at the same time (synchrony), in the opposite way at the same time (anti-phase synchrony), or not at all. While the performing this task, the researcher would read a list of purportedly distracting words that participants were told to ignore. Participants in the synchronous condition were the most accurate in identifying a photograph of the experimenter’s face (rather than face-morphed versions), and in remembering the words they were told to ignore, with the anti-synchrony performing second-highest. These results indicate that the act of synchrony with another person, in-phase or anti-phase, can improve memory for facial and verbal information associated with that person (Macrae, Duffy, Miles, & Lawrence, 2008). Along with a sense of team membership and increased affinity for another, synchrony may result in boosted attention for that agent, perhaps regardless of whether the agent is said to be human or machine.

In another study of behavioral matching and alignment, participants displayed synchrony for facial expressions and gestures while performing a route-communication task (Louwerse, Dale, Bard, & Jeuniaux, 2012). Synchrony was shown to increase as the task became more difficult, which could indicate that synchrony is an important element of functional communication. There is also some evidence that people may align with computer interlocutors as much as with other humans, if not more so, perhaps to make their speech more understandable to an agent with a limited vocabulary (Branigan et al., 2010).

Certain motion synchrony effects have also been demonstrated in studies of human-robot interaction. In one study, a dancing robot named Keepon was put on display with a hidden camera and a sign prompting passersby to dance (Michalowski, Sabanovic, & Kozima, 2007). There was music playing, but Keepon was programmed to respond only to the movement of the dancer in front of it. This means that Keepon was not always moving in-time with the music, as a result of sometimes having its focus on a participant who had not yet started to dance. Participants responded more positively to Keepon when the robot was dancing in synchrony with the music, and exhibited a number of corrective behaviors when Keepon was out-of-sync (moving in an exaggerated manner, touching the robot and forcing it to move in tempo, or standing still and then dancing when the robot started to move in tempo). While observational in nature, this study offers preliminary evidence indicating that when people are given music to move in time with, they desire synchronous interactions with machines.

In a follow-up study, researchers tried to make it possible for Keepon to lead or follow in a dance-like interaction, using input from a Wii remote and balance-board in order to lock onto a child's motion and synchronize with it. Keepon was most successful in encouraging dancer retention when synchronizing with music rather than with the children themselves (Michalowski, Simmons, & Kozima, 2009). That said, when the robot was designated the leader of the dance, children did synchronize with the robot even when it was moving out-of-sync with the music (and children spent more time in-sync with music while following Keepon's lead than when leading the dance themselves, which might indicate some flaw in Keepon's ability to detect the motion of participants). Due to the nature of the design, this study is hard to generalize outside of contexts involving children and dance. Still, there are signs that even lifeless and mindless machines may encourage synchrony with their users.

In one study I collaborated on, we used induced motion synchrony as a means of boosting the rate of sarcasm production within a conversational exchange (Hammond, D'Arcey, Larson & Fox Tree, *in preparation*). In this study, pairs of participants came in and engaged in 3 different movement activities (Simon Says, Ball Passing, and Emotional Mirroring tasks), either while facing each other or back-to-back. Next, these dyads engaged in a ten-minute conversation about badly dressed celebrities. Finally, the dyads marked instances of sarcasm within a video of their conversation as it was produced by themselves or by their partner. The collaborative movement condition (facing partner) resulted in higher self-report of sarcasm than the non-collaborative movement (back-to-back), and participants reported a higher rate of sarcasm use by their partner when they reported higher rate of sarcasm in themselves. We conclude that collaborative movement can boost self-reported sarcasm, perhaps by increasing rapport and prosocial behavior through a synchrony exercise.

In a series of three experiments, I seek to demonstrate that behavioral matching techniques that have been used to produce a sense of social rapport can also be used to similar effect in human-computer and human-robot interactions. If machines can be humanized as agents by reframing backstories and through collaborative synchronous activities, then it might be possible to rehumanize human outgroup members in a similar way.

Current Studies

I speculate that motion synchrony exercises could indeed be used to reduce the self-other distinction between people of different groups, perhaps even between humans and machines. Having people engage in activities that encourage them to converge with another entity could bring them to humanize the human/machine agent they are interacting with to a greater extent. If this can be demonstrated in human-computer and human-robot interaction, then perhaps it could also be used to benefit other ingroup/outgroup relations and reduce social stigma. In three experiments, I will investigate how to reshape first-impressions by engaging participants in motion synchrony interactions with a non-human agent (either human or machine).

In Experiment 1, participants trained a couple of two-dimensional, stationary, cartoon robots to make accurate clinical diagnoses (one robot designed to aid therapists and the other to replace therapists, both normed in a Mechanical Turk survey study). Next, participants who volunteered for an additional training task mimicked the first robot they encountered by repeating the motion they saw on the screen. Finally, participants rated the robots along a number of humanization traits in a survey following each task.

In Experiment 2, participants interacted with a robot, communicating via button-press in a choose-your-own-adventure conversation tree design before and after a neck-stretching motion synchrony task. After this interactive task, participants filled out a survey on humanization as well as memory for the robot. The type of responses selected during the interaction with the robot were compared between groups. We expected that participants would give more positive and warm responses following the synchrony task than participants who performed the neck-stretching activity without synchrony.

In Experiment 3, participants communicated with an addressee who is said to be either a machine learning algorithm or another student, who communicated with the participant via text on a contentious subject of debate. The addressee was set to always disagree with the participant, so as to encourage a sense that the addressee belongs to an outgroup. After a short survey, participants interacted in a rhythm-based task with the same agent (with an egg shaker) before completing a final survey to measure perceived humanization and persuasion of the virtual agent.

Hypotheses tested included: (1) humanization measures – both human nature and uniqueness traits – can be applied to the study of human-robot interactions; (2) motion synchrony and behavioral matching can lead to more positive interactions and higher humanization of a supposedly machine or human agent; (3) participants will humanize a supposedly human agent more than a machine agent, but collaborative synchrony can increase humanization for humans and machine agents alike.

Norming RoboShrink and TheraBot – "Robot Design Survey"

In order to ascertain whether the visual stimuli used in Study 1 were appropriately matched to the name RoboShrink and TheraBot, I conducted a Mechanical Turk norming experiment, asking participants to fill out a quick survey and rate either the names or the pictures along a number of scales.

Method

Participants. Participants were 20 United States Citizens who listed English as their native language, collected via Amazon's Mechanical Turk ("Master Turkers" only) who each received fifty cents for participating.

Materials. The participants also were asked to complete a Google Forms surveys, making use of the *shuffle* feature within each page to randomly change the order the questions were presented in for each participant. They were asked questions pertaining to the perceived characteristics of the two robot designs, one triangular and one round (Figures 1a and 1b respectively) and whether "RoboShrink" and "TheraBot" names were better-suited to one or the other design. As the artist, I intended for the triangular robot design to suit the more serious therapist replacement role, and the round robot design to suit the cute and helpful therapist assistant role.

Design. This was an exploratory norming study, with all participants in one group, answering the same survey questions.

Procedure. Master Turkers saw this task on Amazon's Mechanical Turk website and choose to participate or not on a voluntary basis. After completion of the survey, participants received a key code that they returned to Mechanical Turk to enter and receive payment. The survey itself took an estimated five minutes to complete.

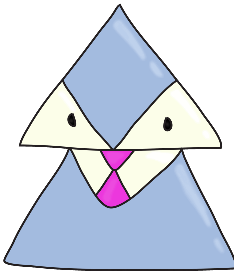
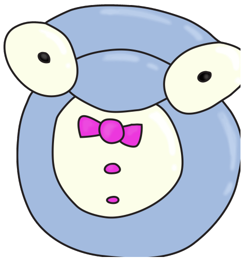
Figure 1a.	**Figure 1b.**
	

Figure 1. The two robot stimuli, (a) RoboShrink and (b) TheraBot.

Results

Robot names. Of the twenty participants who participated in the survey, 13 (65%) preferred the name "RoboShrink" for the triangular robot design, while 11 (55%) of participants preferred the name "TheraBot" for the round robot design.

Replacement or helper. When asked which robot would better serve as a therapist replacement, 10 (50%) selected the image of the triangular robot, 9 (45%) selected the round robot, and 1 (5%) indicated "possible find another design" as a write-in option. When asked which robot would better serve as a therapist helper-bot, 16 (80%) selected the image of the round robot, with 4 (20%) selecting the triangular robot.

Robot qualities. Participants were asked to respond on a Likert scale of 1-7, with 1 being "Not at all" and 7 being "Definitely," on whether a given robot name sounded threatening, friendly, trustworthy, and whether they would feel comfortable with a therapist of that name. I conducted a series of paired-samples *t*-tests for each quality, and found one significant difference between groups for the name sounding threatening. The other differences were not significant, but I report them here as well. RoboShrink was rated significantly more threatening (*M* = 3.3, *SD* = 2.03) than TheraBot (*M* = 2.1, *SD* = 1.48), *t*(19) = 2.77, *p* = .012. RoboShrink was trending less friendly (*M* = 3.85, *SD* = 1.6) than TheraBot (*M* = 4.8, *SD* = 1.7), though not quite significant, *t*(19) = 1.94, *p* = .067. RoboShrink was trending less trustworthy (*M* = 4.15, *SD* = 1.98) than TheraBot (*M* = 4.3, *SD* = 1.69), though also not significant, *t*(19) = 0.29, *p* = .776. Finally, participants tended to rate themselves less comfortable going to RoboShrink (*M* = 3.45, *SD* = 1.64) as compared to TheraBot (*M* = 4.25, *SD* = 2.02), though also not significant, *t*(19) = 1.75, *p* = .096.

Robot gender. For the triangular robot design, 17 (85%) participants described its gender as “male,” as opposed to 2 (10%) saying “female” or 1 (5%) saying the triangular robot was not gendered. For the round robot design, only 9 (45%) participants described the gender as “male,” as opposed to 8 (40%) selecting “not gendered,” and 3 (15%) selecting “female.”

Overall preference. When asked which design they liked better overall, 12 (60%) said round robot, 7 (35%) said triangular robot, and 1 (5%) wrote-in “dislike both.”

Situational preference. Participants responded to a series of forced-choice questions regarding their preference for the name “RoboShrink” or

“TheraBot” for different situations. When asked which robot name appealed to them more when paying for chat-bot therapy, 11 participants (55%) selected “TheraBot.” When asked which computer tool they would pay for as a therapist to help them in making diagnoses, 14 (70%) selected “TheraBot.” When recommending a chat-bot service to a friend with depression, 14 (70%) selected “TheraBot.” As a programmer, 15 (75%) said they would name their chat-bot “TheraBot” if given the choice.

Discussion

I designed the two robot images specifically to encourage participants to pay closer attention to the main experimental manipulation (the backstory), with one designed as a therapist replacement, and the other designed as a therapist helper-bot. Participants expressed a slight preference for the triangular bot as a replacement bot (50% as compared to 45%) and a clear preference for the round robot design as a helper bot (80%). A majority of participants preferred the name “RoboShrink” for the triangular bot design (65%), while a slight majority preferred “TheraBot” for the round robot design (55%).

Unbeknownst to participants, the lead researcher designed the triangular robot with the name “RoboShrink” and replacement-bot in mind, and the round robot with the name “TheraBot” and the helper-bot in mind. The angular design was meant to elicit a sense of seriousness and maturity, while the round design was meant to elicit a sense of cuteness and childishness. When asking participants their initial impressions of each design, the comments indeed reflect that the intended purpose came across. Comparing the triangle and round designs, one participant described them as, “A bureaucrat” and “An entertaining balloon animal” respectively. Another said, “It appears a bit sharp and kind of cutting. Comes across a bit hard and a little angry” and “It makes me think of a penguin and it's kind of cute and cuddly” Another participant wrote, “A professor someone that is wise and serious” and “This one seems juvenile like I wouldn't be able to take it seriously” Finally, one participant compared them to other animals, saying, “A snake. Cold and potentially mean” and “A penguin. Fun and friendly.”

Participants in this norming task provide support for the use of a triangular RoboShrink as the serious and mature design, and a round TheraBot as the cute and childlike design. Only a slight majority expressed an overall preference for TheraBot over RoboShrink, however (60%). It is also interesting to note that RoboShrink was viewed as having a gender (with most participants, 85%, describing it as male) more so than TheraBot (with only slightly more describing the robot as male, 45%, as compared to not-gendered, 40%). It is unclear whether these gender perceptions were driven by the perception of other characteristics the robots seemed to exhibit, which fell into categorical stereotypes of gender, or whether it was from something more integral to the design.

Experiment 1 – "Getting The Help They Need": Preferences for a Helper or Replacement Robot Therapist

With the robot designs tested and confirmed for their roles as helper or replacement robot therapists, I designed Experiment 1 as the first in a series of experiments to demonstrate that humanization scales (previously used only in studies of humans) and behavioral matching tasks (mimicry) can both be used in the study of human-robot interactions. I predicted that participants would prefer to interact with and train the helper bot, TheraBot, over the replacement bot, RoboShrink, in line with research I had formerly collected from the same participant pool that indicated a strong preference for human over chat-bot therapists (Larson & Fox Tree, *in review*). I also predicted that the voluntary mimicry task would result in a boost for the humanization scores of both RoboShrink and TheraBot, as it would increase a sense of team-membership with the machine (although perhaps not as effectively as a motion synchrony task).

Method

Participants. Participants were 64 undergraduate students at the University of California, Santa Cruz, who each received course credit for participating. Of these, 20 participants volunteered and participated in an additional mimicry task. A total of 24 participants answered questions on the post-study survey that indicated the participant had not read and fully understood the stimuli (of these, 9 were volunteers for the additional mimicry task), and were consequently excluded from all but the *impression change after mimicry* analysis. For that set of comparison, participants who did not complete the additional training task were excluded, but participants who failed one or more of the manipulation checks were re-included, as we were interested in changes in impressions before and after the mimicry task.

Materials. The first experimental task was coded in Python 3.0, using the AppJar GUI Library for an interactive presentation of stimuli, as well as Pandas for saving output to a CSV file for each participant's in-task reaction time and accuracy data. The task itself was run on Mac OS X (version 10.6.8). This task also included two versions of two robot design images for "RoboShrink" and "TheraBot" (triangular design and round design, respectively) either as a floating image, or as an image within a therapist office scene, either sitting on the therapist chair (Figure 2a), or beside it (Figure 2b), respectively.

The text-based experimental stimuli took the form of two separate backstories for the robots (Table 1), as well as case stories that participants used to select a diagnosis from three multiple choice options. These case studies were collected and adapted by research assistants in the lab from a variety of sources (Cooper, 2011; Gillig, 2009; Lane, 2017; Manning, 1999; Rolls, 2015; Shah & Nakamura, 2010; Spitzer, Skodol, & Gibbon, 2002; Trull & Prinstein, 2013). The additional mimicry task included face-masked stimuli from a library of GIFs on American Sign Language found online at *http://www.lifeprint.com/asl101/gifs-animated/*. Participants of the optional task were recorded using QuickTime Player (version 10.4) and Logitech HD Pro Webcam C920. The participants also were asked to complete either 1 or 2 Google Forms surveys, making use of the *shuffle* feature within each page to randomly change the order the questions were presented in for each participant.

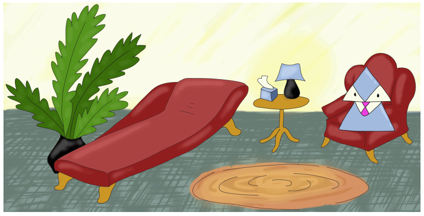
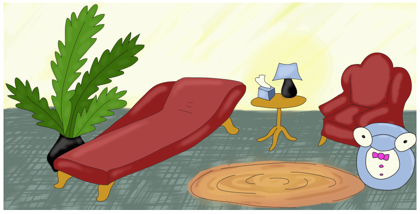
Figure 2a.	**Figure 2b.**
	

Figure 2. The two robot stimuli in full scene, (a) RoboShrink and (b) TheraBot.

Table 1.	
RoboShrink	“We are testing another computer algorithm called RoboShrink. It is designed to REPLACE human therapists with a cheaper and more convenient option for treatment.”
TheraBot	“We are testing a computer algorithm called TheraBot. It is designed to HELP therapists in treating their patients, not to replace human therapists altogether.”

Table 1. The two sets of instructions, for RoboShrink and TheraBot.

Design. I used a 2 (“RoboShrink” Replacement or “TheraBot” Helper Backstory) x 2 (Motion Synchrony or Non-Motion Task) within-subjects design. Each participant saw the same robot designs, but half of participants saw RoboShrink first, and half saw TheraBot first (to control for possible order or practice effects). Participants all saw case studies 1-10 for the first robot in shuffled order, followed by case studies 11-20 for the second robot in shuffled order. Each case study was assigned a specific set of multiple choice options (designed to be correct, almost-correct, and incorrect) organized in such a way that both robots would answer the same number of times correctly (3), almost-correctly (4), and incorrectly (3), while selecting each option (1, 2, or 3) the same number of times, (3, 4, and 3 respectively).

After finishing the first task, participants were asked to participate in an optional video-recorded motion task to help train the first robot they read about. The accuracy and reaction time within the first task, the ratings of the two robots along a humanization scale, as well as a direct comparison of the robots and a rating for preference of one over the other will all act as dependent variables of interest. Additionally, I will look to see if the humanization rating of the first robot changes from the first survey to the second (for those who choose to participate in an optional sign language behavior-matching mimicry task). I also asked the participant to rate their interest in clinical psychology, as well as experience with computers.

Procedure. Participants first completed a clinical case study diagnosis task. This task was composed of the following: Introduction to first robot, 10 case study multiple choice trials, introduction to second robot, 10 more case studies, and a final thank-you. The introduction screens described either RoboShrink or TheraBot, as well as the picture of the robot in a therapist's office (Figure 2). These two introductory instruction slides appeared for 20 seconds before the participant was allowed to press space to continue. After viewing the first instruction slide where they were introduced to the first robot, participants saw the first case study slide, with a description of a fictional character to diagnose and 3 multiple choice options they could select between for diagnosis.

First the robot made a selection out of the 3 options by underlining the text, with the picture of the robot alongside the options. The participant could then confirm the robot's choice or make a different one by pressing 1, 2, or 3, and their response was bolded. They could change their response an unlimited number of times before pressing space to continue to the next case story. Participants were not told that they would be seeing more than one robot, but they would be introduced to the second robot after they completed the 10th case study multiple choice question. Trials 11-20 proceeded in the same way as 1-10, but with different case studies (the stories were shuffled within these two blocks). The final screen participants saw thanked them and asked for their feedback in an upcoming survey.

After they completed this survey, the research assistant said, "You're all done with the experiment, but would you like to stay and help us improve <*first-robot-name*> in a short video-recorded task?" If they decided to volunteer for the optional task, they signed a second consent form specifically for the video release, and were escorted to a different computer room that was set up with the video-recording equipment and a Google Slide presentation including American Sign Language movements that they were asked to imitate, in order to help the robot learn to "interact with people in real-time and understand body language," according to their instructions. They were reminded that they could leave at any time. The research assistant maked a note of their start and end times, as well as which slide they stopped on. The participants filled out a short final exit survey with questions similar to those on the previous survey, but this time only for the robot they has just helped train, and were thanked for their time. All participants received course credit for participation.

Results

Subjective measures. Of primary interest in the design of this experiment were the subjective ratings of humanization, and subscales (Human Nature and Human Uniqueness), for both robot designs, as well as before and after the optional mimicry task.

Initial robot preferences. I conducted a repeated-measures ANOVA (to control for possible order effects of robot presentation) and found that there were no significant differences in the humanization of RoboShrink as compared to TheraBot on Survey 1, *F*(1,38) = 1.23, *p* = .274.

Excluding participants who did not identify their preferred robot, I conducted a one-way ANOVA to look at the combined Human Nature and Human Uniqueness scores (grouped by positive and negative) for RoboShrink and TheraBot to see if higher preference was related to more humanization of a given robot. Participants who expressed an initial preference for Roboshrink rated it more highly for Human Nature traits (*M* = 3.9, *SD* = 1.76) compared to those who preferred TheraBot (*M* = 2.56, *SD* = 0.96), *F*(1,32) = 8.02, *p* = .008. Those who preferred RoboShrink also rated it more highly for Human Uniqueness traits (*M* = 4.63, *SD* = 1.48) than those who preferred TheraBot (*M* =3.5, *SD* = 1.17), *F*(1,32) = 5, *p* = .032. Although there were no statistically significant differences between groups on ratings of negative qualities for RoboShrink, the combined Human Nature and Human Uniqueness scores followed the same pattern as the aforementioned results. For Human Nature traits, RoboShrink was rated more highly overall by RoboShrink than TheraBot fans (*M* = 4.91, *SD* = 0.89 & *M* = 4.15, *SD* = 0.54, respectively), *F*(1,32) = 8.81, *p* = .006. For Human Uniqueness traits, RoboShrink was rated more highly overall by RoboShrink than TheraBot fans (*M* = 4.96, *SD* = 0.86 & *M* = 3.9, *SD* = 0.74, respectively), *F*(1,32) = 11.66, *p* = .002. This means that overall, the combined Humanization score for RoboShrink was higher among those who preferred RoboShrink (*M* = 4.94, *SD* = 0.8) than those who preferred TheraBot (*M* = 4.03, *SD* = 0.57), *F*(1,32) = 13.01, *p*

= .001. There were no such group differences for humanization ratings of TheraBot.

For RoboShrink, within-subjects positive scores for Human Nature and Human Uniqueness items were highly correlated, for both positive items (*r* = .755, *p* < .001), negative items (*r* = .53, *p* < .001), and overall combined humanization scores (*r* = .69, *p* < .001), while Human Nature and Human Uniqueness scores differed significantly for all but the combined score metric. A paired-samples t-test revealed that for positively-valenced items, participants rated RoboShrink more highly for positive Human Uniqueness traits (*M* = 3.85, *SD* = 1.26) than Human Nature traits (*M* = 2.92, *SD* = 1.35), *t*(39) = 6.32, *p* < .001. For negatively-valenced items, participants still rated RoboShrink more highly for negative Human Uniqueness traits (*M* = 3.5, *SD* = 1.33) than Human Nature traits (*M* = 2.19, *SD* = 1.18), *t*(39) = 6.70, *p* < .001. Overall, participants gave RoboShrink a positively-valenced average score that trended higher for Human Nature (*M* = 4.37, *SD* = 0.69) than Human Uniqueness (*M* = 4.17, *SD* = 0.88) combined scores, though not significantly, *t*(39) = 1.92, *p* = .062.

For TheraBot, within-subjects positive scores for Human Nature and Human Uniqueness items were highly correlated, for positive items (*r* = .61, *p* < .001), and overall combined scores (*r* = .422, *p* = .007), though not for negative items, while Human Nature and Human Uniqueness scores differed significantly for all but the combined score metric. A paired-samples t-test revealed that for positively-valenced items, participants rated TheraBot more highly for positive Human Uniqueness traits (*M* = 4, *SD* = 0.99) than Human Nature traits (*M* = 3.06, *SD* = 1.14), *t*(39) = 6.24, *p* < .001. For negatively-valenced items, participants still rated TheraBot more highly for negative Human Uniqueness traits (*M* = 3.16, *SD* = 1.2) than Human Nature traits (*M* = 2.28, *SD* = 1.08), *t*(39) = 4.01, *p* < .001. There were no significant differences for TheraBot's combined scores on Human Nature versus Human Uniqueness.

The combined humanization score for RoboShrink was inversely correlated with the preference score for TheraBot (from a combination of responses to scenarios where one robot would be better than the other), *r* = -.35, *p* = .027. However, while humanization of TheraBot was positively trending with the combined preference score for TheraBot, it was not significantly correlated, *r* = .29, *p* = .071. A one-way ANOVA analyses did not indicate that there were significant differences in the humanization of a robot based on perceived gender-similarity between participant and robot.

Impression change after mimicry. I examined whether humanization impressions on the robot that participants were asked to train improved following the optional mimicry task. These tests revealed that for participants who trained RoboShrink, there was a trending correlation between the first time they rated RoboShrink on positive (*r* = .55, *p* = .067), and high correlation on negative (*r* = .62, *p* = .031) Human Nature items. These participants who trained RoboShrink also had a high correlation between the first and second time they rated RoboShrink on positive (*r* = .69, *p* = .012) and negative (*r* = .69, *p* = .013) Human Uniqueness items. There were no high correlations within-subjects for TheraBot training and humanization, however.

That said, there were significant differences in overall humanization scores of the robot that participants opted to train in the additional mimicry task. A one-way ANOVA reveals that participants who trained TheraBot in the post-experimental mimicry task also gave the bot they trained higher overall scores for Human Nature (*M* = 5.2, *SD* = 0.66) than did participants who trained RoboShrink (*M* = 4.28, *SD* = 0.42), *F*(1,18) = 14.9, *p* = .001. Participants who trained TheraBot in the post-experimental mimicry task also gave the bot they trained higher overall scores for Human Uniqueness (*M* = 4.96, *SD* = 0.5) than did participants who trained RoboShrink (*M* = 4.23, *SD* = 0.88), *F*(1,18) = 4.5, *p* = .048. Moreover, participants who trained TheraBot in the post-experimental mimicry task also gave the bot they trained higher overall scores for combined Humanization scales (*M* = 5.08, *SD* = 0.52) than did participants who trained RoboShrink (*M* = 4.25, *SD* = 0.62), *F*(1,18) = 9.71, *p* = .006.

Behavioral measures. Aside from subjective responses in the post-experimental survey, other more behavioral measures were compared. These included oder effects, reaction time, accuracy in the case diagnosis task, and participation in the mimicry task.

Order effects. Participants took more time (in milliseconds) to complete the first ten case studies (*M* = 353.61, *SD* = 89.94) than the second set of ten case studies (*M* = 314.24, *SD* = 118.21), *t*(39) = 3.51, *p* = .001. This demonstrates a learning and possibly practice effect. In addition, Participants tended to be less accurate on the first ten case studies (*M* = 17.23, *SD* = 2.28) than on the second ten case studies (*M* = 18.38, *SD* = 2.28), *t*(39) = 2,7, *p* = .01. The time a participant took to complete Block 1 and Block 2 was highly correlated (*r* = .801, *p* < .001), but a participant's accuracy on Block 1 and Block 2 only tended to be correlated (*r* = .303, *p* = .057).

People had a strong preference for the first robot they encountered. Those who saw TheraBot first were more likely to give RoboShrink higher ratings for negative Human Nature traits (*M* = 2.61, *SD* = 1.26) compared to those who saw RoboShrink first (*M* = 1.77, *SD* = 0.96), *t*(38) = 2.37, *p* = .023. Those who saw TheraBot first tended to give TheraBot, by contrast, higher ratings for positive Human Nature traits (*M* = 3.37, *SD* = 1.03) than those who saw RoboShrink first (*M* = 2.74, *SD* = 1.19), *t*(38) = 1.79, *p* = .082. Those who saw TheraBot first also tended to give TheraBot higher ratings for positive Human Uniqueness traits (*M* = 4.28, *SD* = 0.74) than those who saw RoboShrink first (*M* = 3.71, *SD* = 1.13), *t*(38) = 1.89, *p* = .067. There was also a tendency for those who saw TheraBot first to give RoboShrink higher ratings for negative Human Uniqueness traits (*M* = 3.88, *SD* = 1.14) than those who saw RoboShrink first (*M* = 3.11, *SD* = 1.42), *t*(38) = 1.89, *p* = .066.

Reaction time. A paired-samples *t*-test revealed that participants responded similarly on the case study task regardless of which robot they were interacting with, *t*(39) = 0.9, *p* = .377, with high correlation of total reaction time (*r* = .67, *p* < .001) for RoboShrink and TheraBot trials. No significant differences between robot trials were found for reaction time or accuracy.

Accuracy in task. I conducted a one-way ANOVA to see if response accuracy in the clinical trial task for case identification differed between the participant's preferred bot (if participants were more likely to choose the distractor answers that one robot selected over the other), but found no significant differences after removing participants who selected something other than RoboShrink or TheraBot as their preferred bot.

Participation in optional task. An independent samples *t*-test revealed that participants who preferred RoboShrink and passed the manipulation checks spent significantly more time gesturing in the optional training task (*M* = 20.83, *SD* = 4.12) than participants who preferred TheraBot (*M* = 12.6, *SD* = 6.91), *t*(9) = 2.46, *p* = .036. When looking for possible demographic differences between the participants who opted to participate in the additional mimicry task or not, an independent samples *t*-test indicated a pattern that went against what I had previously predicted. Participants who opted to train RoboShrink in the post-experimental mimicry task were more likely to have rated themselves as interested in pursuing a career in therapy (*M* = 6.25, *SD* = 0.87) than those who opted to train TheraBot (*M* = 5, *SD* = 1.2), *t*(18) = 2.72, *p* = .014.

Discussion

Experiment 1 included not only humanization measures, but also several behavioral measures, designed to assess the importance of visual and backstory cues on reactions to the two robot designs.

Humanization measures. The humanization measures used here are typically reserved for describing the perceived characteristics of humans (Crawford, Modri, & Motyl, 2013; Haslam, 2006) but, as demonstrated here, can also be applied to robots. A priori, it seemed reasonable to expect that the humanization scale may not work the same way for humans as machines, so it was important to conduct this first experiment and see what kinds of ratings the robots tend to get.

In fact, people were happy to apply humanization judgements to robots. A couple participants picked the middle number in the scale, “4,” for every item in the humanization scale ($n = 2$), but this was uncommon (3.13% of the sample). In addition, judgements of one robot influenced judgements of another robot. People who saw TheraBot first gave RoboShrink a higher score for both negative and positive Human Nature and Human Uniqueness traits, compared to those who saw RoboShrink first.

Nonetheless, the humanization scale may function differently for humans than for machines – even negative human nature and uniqueness traits are still *humanization* traits. This means that instead of looking at the scale as being a human–nonhuman rating system, it is more informative to look at it as a way of measuring whether the robot is (1) exhibiting more features of a living, conscious being and (2) whether those features are seen as prosocial or antisocial.

Mimicry with machines. In addition, behavioral matching tasks (here, mimicry) typically used in human-human contexts can be used in human-computer or human-robot contexts in order to improve relations between a human and machine. The humanization ratings for RoboShrink were highly correlated between the first rating, and the post-mimicry rating, but this wasn’t the case for TheraBot humanization ratings. While there were no clear differences in humanization ratings for RoboShrink and TheraBot in the first survey, humanization ratings for TheraBot were higher than those of RoboShrink following the optional mimicry task.

After participating in the optional training task, and looking at between-subjects differences in ratings for the bot they helped train, we found that participants who trained TheraBot *did* rate it higher on the humanization scales than those who trained and rated RoboShrink after the mimicry task. This result indicates that differences in backstories and appearances can lead to some differences in the way people rate the machines on these scales if they are asked to perform a behavioral matching task with the robot. It is possible that mimicry can improve humanization ratings for cute robots, but not for serious or emotionally distant robots.

Participant accuracy. Of course, this experiment was not without its challenges. A pilot version of this experiment indicated that participants had difficulty accurately remembering the backstories for robots (this experiment originally had a between-subjects design where half of participants read the "replace" and the other half read the "help" backstories for RoboShink). I opted for a within-subjects design in hopes that participants would be able to directly compare the backstory of one robot to another, and yet 24 of the 64 participants failed at least one of the manipulation check questions (two open-ended questions first, followed by two multiple-choice questions in the post-experimental survey). While I had expected that the participants who chose to volunteer would be the more attentive and diligent participants (to put in extra time that was not required of them), I in fact found the opposite to be the case. The fail rate for the participants who didn’t volunteer was about 34%, while it was 45% for those who did choose to volunteer for the additional video-recorded mimicry task. So even though the participants all saw the word “REPLACE” or “HELP” in all-caps letters on the instruction screens, and saw both robots (who had been normed for the best fit of design to function), over a third of participants had difficulty remembering which was which.

Looking more closely, the participants who failed on one test question often failed on multiple questions, with a total of 29 (45.3%) participants failing at least one manipulation check question on the purpose of RoboShrink, and 15 (23.4%) failing at least one question on the purpose of Therabot. As this is the same pattern observed in the pilot version of this task, it is possible that the huge fail rate is possibly resulting from a bias to think of robots as assistants rather than replacements for humans performing emotional work, despite the fact that such robots are out there and becoming more popular all the time (Fitzpatrick et al., 2017).

As a potential counter-argument to the notion that *computers are making us stupid*, we found that participants were more likely to be influenced by the bad choices of a robot that they trained in Block 1 of the clinical diagnosis task as compared to Block 2. Perhaps this is an encouraging finding, that people may be influenced by a machine, but with increased exposure to a task and to the robots, will begin to trust their own instincts more. In any case, the robot’s incorrect choices certainly don’t seem to impede rational decision-making abilities within the scope of this clinical diagnosis task, once a participant became acclimated to it. While there may be some fears that people put too much trust in machines in a bad way, this can potentially be remedied through increased exposure to different robots.

Individual differences. There seems to be something special about participants who selected RoboShrink as their preferred robot. They were more easily guided to make bad choices by their preferred robot, and humanize their preferred bot more than those who preferred TheraBot, and also spending more time on the optional mimicry task when they chose to volunteer. Completely counter to my prediction, those who opted into training RoboShrink also reported an interest in pursuing a career in therapy more so than those who opted to train TheraBot. From my previous research on human-robot interaction opinions, participants had indicated a strong dislike for the idea of consulting with a chatbot therapist as compared to a human therapist over chat (Larson & Fox Tree, *in review*). If participants who report an interest in therapy as a career opt to train RoboShrink in greater numbers than TheraBot, it seems like they are working to put themselves out of a job (since RoboShrink was designed to replace therapists, while TheraBot was designed to be a therapist's assistant. Or perhaps they are motivated by something truly altruistic: they may be more interested in working to make automation of the workforce possible, in order to benefit those who do not otherwise have access to affordable therapy (since TheraBot would supposedly be used in conjunction with existing therapeutic methods).

Need for physical robots. Experiment 1 provided some useful insights into how people develop and adapt impressions of non-living machine agents, but wasn’t without limitations. Firstly, participants had some difficulty remembering the backstories of the agents, perhaps due to the within-subjects design. It’s possible that a between-subjects design could result in less confusion over the framing of a machine agent, although previous research in the lab suggests that perhaps participants have a hard time grasping the idea that machine agents can be capable of replacing humans in general (Larson & Fox Tree, *in review*). Secondly, I was unable to employ real, physical agents in this task. This means that I was unable to create a true synchrony motion-based task for participants to perform, and needed to instead rely on the use of GIFs for the robot’s purported sign language gestures. There is also reason to believe that the presence of a physical robot in itself can result in a greater degree of situational empathy for the machine (Seo, Geiskkovitch, Nakane, King, & Young, 2015). The upcoming studies will provide useful information regarding the importance of a robot or computer agent’s behavior or human-status, and whether these features will lead to differing benefits from motion synchrony and collaborative interventions. The next experiment makes use of physical robots instead of simple drawings.

Experiment 2 – "Engaging with a Robot":

Interactions with a Robot Before and After a Motion Task

In Experiment 2, people interacted with a machine that bore the likeness of a prototypical robot while performing a motion synchrony task, which is more involved than the simple non-synchronous motion dictation task in the first experiment. As in Experiment 1, participants were asked to assess humanization and other characteristics, and I predicted that the motion synchrony task would lead to more positive ratings along these scales as compared to a non-synchrony task. This second experiment also involved a new set of measures for participants' recollection of the robot's visual and verbal features, with the prediction that synchrony would boost attention on the agent, and therefore memory of the robot agent (Macrae et al., 2008).

Method

Participants. Participants were 161 undergraduate students at the University of California, Santa Cruz, who each received course credit for participating. Of the original 212 participants, 51 were excluded from the analysis, due to equipment malfunction (*n* = 47), power outage (*n* = 1), or because they said they were not fluent in English (*n* = 3).

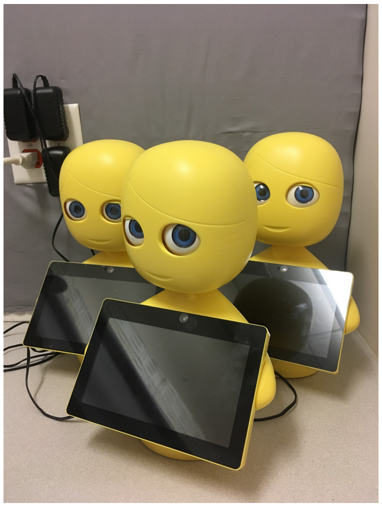
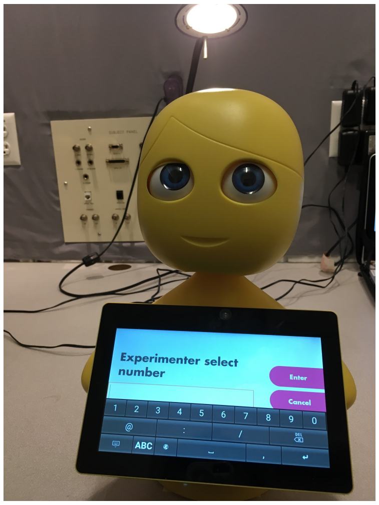
Figure 3a.	**Figure 3b.**
	

Figure 3. Mabu, (a) all three available robots and (b) the participant set-up.

Materials. The robotic agent was one of three Mabu robots from Catalia Health (Figure 3a) that were each programmed using Twine and Lua to perform a series of conversational turns and gestures, as well as stretching motions for the participant to imitate. While the participant interacted with just one of the three robots (Figure 3b), there were two back-up robots in case of malfunction. Each participant communicated with the robot in a series of semi-scripted conversational turns, where they could select one of a number of responses to each of the robot's conversational turns. Survey measures were administered via Google Forms. The robot was positioned at the end of the table in a well-lit room, approximately 8 inches from the participant, to allow for the best view of the participant's face from the perspective of the robot's camera (not used to make recordings, only for face-tracking purposes).

Design. There were two between-subject levels: motion synchrony and non-synchrony. Unlike the previous experiment, this time the embodied agent always had the same story and appearance, and the only noticeable difference was whether it performed the rhythm-based task in synchrony while directing the participant through a series of neck stretches or merely directed the stretches without moving. I compared humanization ratings and memory for the robot between the two groups, but also performed exploratory analyses to see which options the participants selected as they worked through a conversation with the machine, and compared the types of responses they select before and after participating in either a synchronous or non-synchronous motion-based task with the robot. Participants also filled out a brief survey after the task about their memory for what the robot said and looked like, humanization measures, as well as their general opinions about the usefulness of robots as well as their capabilities and whether they would consider buying one sometime in the future.

Procedure. Participants came into the lab, and were then introduced to one of the three effectively identical robots (Figure 3a). They responded to a series of conversational prompts, with the robot always initiating the conversational topic and providing a number of possible responses for the user to select via touch-screen press, where certain responses could indicate a lack of eagerness on the part of the participant and initiate a quitting block. If a participant did not decide to quit during this initial greeting block, they continued on to the motion activity block, where the robot (introduced as Mabu) would either ask them to watch and repeat a set of stretching motions (non-synchrony), or watch and mimic the same stretching motions (synchrony).

Periodically throughout the conversation, participants were prompted by the machine with four possible response options. These were designed to allow for a spectrum of responses, being either essentially positive (agreeable, upbeat, compliant) or negative (disagreeable, a downer, noncompliant) in meaning and warm (empathizing, kind, polite) or cold (un-empathizing, gruff, impolite) in tone. This meant that there were 4 options for the participant on each screen: Positive–Warm, Positive–Cold, Negative–Warm, and Negative–Cold (in that order from top to bottom).

Following the motion activity were two short closing blocks, giving the participant an opportunity to provide some feedback to Mabu's face. Finally, the participant was escorted to a separate room to fill out another short survey, including a long-answer section on possible improvements for better interactions with machines in future iterations of the design, as well as memory for the robot.

Results

Participants in the two groups were compared on overall humanization, memory for the robot, and in-task response valence.

Humanization of the robot. While I expected to find more positive humanization of the agent by participants who participated in the motion synchrony task than those in the non-synchrony task (due to a sense of increased in-group membership), no significant group differences were observed. Still, while the groups did not significantly differ in overall humanization (positive minus negative scores) or any other humanization subscale, the minor differences between synchrony and non-synchrony groups did trend in the predicted direction. The synchrony group rated the robot slightly higher on total average humanization, positive and negative traits combined (*M* = 3.48, *SD* = 0.59) than the non-synchrony group (*M* = 3.38, *SD* = 0.51), *t*(159) = 1.13, *p* = .26, *95%CI* = [-0.07, 0.27]. The synchrony group also rated the robot slightly higher on the positively valenced humanization traits (*M* = 2.84, *SD* = 1.28) than the non-synchrony group (*M* = 2.63, *SD* = 1.44), *t*(159) = 0.99, *p* = .33, *95%CI* = [-0.21, 0.64].

Memory for the robot. While I predicted that memory for the robot's features would be better following the synchrony task as compared to the non-synchrony task, I did not observe any significant difference in memory between the groups. Out of all participants, 59.6% correctly wrote-in the robot's name in the post-experimental survey (only correct spelling of "Mabu" was accepted), and 27.3% of participants correctly identified the robot's stated occupation ("Wellness

Coach”). For the question regarding whether the robot had hands, 47.8% said “yes,” 43.5% said “no,” and 8.7% wrote in some other response. Opinions were less divided on whether the robot had feet, with 18% saying “yes,” 67.1% saying “no,” and 14.9% writing in another response. The write-in responses made it clear that many participants were unsure of what would qualify as “hands” and “feet” and were forced to come up with their own definition.

Response valence. I conducted a series of independent samples *t*-tests, looking at participant response data within the Human–Robot conversation that followed the initial greeting block (before the experimental manipulation). As predicted, participants in the synchronous condition were more likely to select Positive–Warm responses (*M* = 80.03% of a participant's responses, *SD* = 18%) than those in the non-synchronous condition following the stretching activity (*M* = 71.07% of a participant's responses, *SD* = 24.58%), *t*(148.51) = 2.65, *p* = .009, *95%CI* = [0.02, 0.16]. Conversely, participants in the synchronous condition were less likely to select Positive–Cold responses (*M* = 16.6% of a participant's responses, *SD* = 15.6%) than those in the non-synchronous condition following the stretching activity (*M* = 25.17% of a participant's responses, *SD* = 21.55%), *t*(147.69) = 2.9, *p* = .004, *95%CI* = [0.03, 0.14].

Overall, Positive responses were more popular than Negative responses (97.15% as compared to 2.85% of all valenced responses). In fact, a majority 75.78% of participants only selected Positive responses, with only 24.22% of participants opting for at least one Negative response. Positive–Warm responses were the most popular response option from conversation start to finish (74.43%

of responses), with Positive–Cold as the second-most frequent selection (22.73% of responses). Negative–Warm responses were selected a mere 2.43% of the time, and Negative–Cold responses were only selected 0.41% of the time. Considering the low frequency of Negative responses, there were too few responses to make an assessment.

Discussion

Experiment 2 included the same humanization scale used in the first experiment, but also some more exploratory measures, including memory for the robot agent, and response valence in the task. The robots presented some unique challenges, also discussed in the following section.

Humanization of the robot. There were no significant group differences on total humanization scores (positive minus negative traits) or on Human Uniqueness and Human Nature subtotal scores. Previously, in Experiment 1, participants were introduced to a friendly robot assistant for a therapist (TheraBot) and a potentially threatening robot replacement for a therapist (RoboShrink). Participants could then either volunteer to mimic TheraBot or RoboShrink, and humanization scores were measured before and after this mimicry task.

In Experiment 2, the robot was never intentionally designed to pose a threat to the Psychology undergraduate population of study, and I only included a measure of humanization following the rhythm task, so I couldn’t compare within the participants to see if synchrony resulted in a positive change in the score. The humanization scale was included mostly for consistency across experimental measures for eventual comparison between Experiments 1, 2, and 3 (to compare humanization of robot drawings, physical robots, and supposed machine agents versus human agents, respectively). It makes a certain amount of sense that participants would not have cause to dehumanize the robot more in one group than the other in this experiment. That said, I was hopeful that the synchrony task would lead participants to identify the robot more as an in-group member and therefore give the robot a more positive humanization score overall.

In Experiment 3, participants were introduced to a human or algorithm addressee that disagreed with them on a political topic. I expected this to prompt participants to consider their addressee as an out-group member, who might potentially be re-humanized following a synchrony task.

Memory for the agent. In Experiment 2, I hoped to find differences between groups on memory for the robot's features and character details, given prior literature in human–human interaction (Macrae, Duffy, Miles, & Lawrence, 2008). However, the relatively simplistic design of the robot (as compared to a human face) made this difficult. Most participants could accurately recall the robot's color (yellow) and eye color (blue with black pupils), with few guessing other colors. What's more, the questions about whether the robot had hands and feet became a rather more philosophical question than I had anticipated. For example, when asked whether the robot had hands, many participants wrote in a response like: "Not definitive hands I could see, but it was holding the Ipad *[sic]* with arms--not sure if there were 'hands'." and "Yes (sort of) it was able to hold the tablet but I did not necessarily see appendages that resembled hands.” and “I'd day *[sic]* they were more like nubs.”

Because of this, the only two useful measures of memory turned out to be questions on details from the initial greeting block of the conversation, where they learned the robot's name and occupation. But because these details preceded the stretching activity and were written and spoken rather than purely visual in nature, the ability of the participant to remember these details is of limited value, and I did not find that groups differed in how well they could recall these details.

Response valence. The overwhelmingly Positive responses to the robots could potentially result from a desire to be agreeable and polite when interacting with machines, as demonstrated in much of the literature on machines as social agents (Nass, Moon, & Carney, 1999). Regardless of whether they interacted with the robot in synchrony or not, participants didn’t want to be rude to a robot by responding with “I'm not sure I got it right” (Negative–Warm) or “Eh, I didn't really try” (Negative–Cold), and preferred to select responses like “I think I did it correctly” (Positive–Warm) and “I guess I did alright” (Positive–Cold) instead.

While Negative responses were designed to be disagreeable, dissenting, and non-compliant, Positive responses were designed to be agreeable, upbeat, and compliant. Because students who choose to participate were doing so on a voluntary basis, it is likely that they felt little cause to be non-compliant or disagreeable. Similarly, Cold responses were designed to be non-empathetic, gruff, and impolite, while Warm responses were designed to be empathetic, kind, and polite. Due to the endearing qualities of the robot agent, participants may have generally felt a inclination to respond warmly to the machine agent. Whether participants chose to respond warmly or coldly to the robot likely said more about their actual feelings about the robot than the positivity or negativity of the response, which was a better indicator of how they were feeling about the experiment as a whole.

The ordering of valenced responses (always Positive–Warm, Positive–Cold, Negative–Warm, Negative–Cold) might have been part of the reason why participants mostly selected Positive over Negative responses. However, the fact that there were group differences in selection of Positive–Warm vs. Positive–Cold indicates that response selection could be influenced by synchronous interaction with the robot. Order of response options alone could not be the driving force in response selection.

Robot caveats. In Experiment 2, physical robots were used in place of two-dimensional pictures, but this resulted in certain limitations. The robots themselves would occasionally act up and need to be temporarily put out of service. Thankfully, with three robots available, we were able to do a quick switch in most cases. The exception to this is when the mechanical problem occurred after the very start of the experiment, after the participant had already been introduced to the robot by name, we were unable to restart the experiment. The robots were capable of eye-tracking, but only at very close proximity.

Additionally, the participant was instructed to sit close to the robot, but this in itself could lead to a certain degree of discomfort (See Walters, Syrdal, Koay, Dautenhahn, & te Boekhorst, 2008). One participant wrote, "at first i thought it was strange why we have to sit so close to the robots but i figured it was for the robot." If the participant moved too far away for a more comfortable position, eye tracking would not work as well. It is perhaps because of this that some participants wrote things like "i never knew if it was actually looking at me or not" in the post-experimental questionnaire. In addition to this problem, some researchers suggest that when robots are perceived as having emotional experiences (moving or looking like a human), this can induce an uncanny valley effect (Gray & Wegner, 2012).

In the next experiment, we will not be using pictures or physical robots, but instead will go back to a study of framing (telling participants they are interacting with a human or machine) and will employ a between-subjects design to reduce possible confusion that might arise from hearing multiple backstories. This final experiment will provide useful insight into how rhythm-based motion activities might improve interpersonal relations between humans or with machine agents who are introduced as members of an outgroup (who disagree with the participant on a political topic).

Experiment 3 – "Rhythm & News":

Bridging Political Differences via Rhythmic Synchrony

In Experiment 3 I worked to extend findings from Experiments 1 and 2 into human-human interaction. In this design, I tested whether presenting an anonymous virtual agent as a human participant or machine-learning algorithm would shape initially-held beliefs about that agent, and whether those initial impressions could be altered with a rhythm-based synchrony game intervention. I used similar humanization measures to assess the participant's impressions of the agent, as well as questions to assess trust and persuasiveness of the agent following the task.

Method

Participants. Participants were 93 undergraduate students at the University of California, Santa Cruz, who each received course credit for participating. Of the original 107 participants, 14 were excluded from the analysis, due to equipment malfunction (*n* = 5), because they failed to follow the rhythm task instructions (*n* = 2), or because they said they were not fluent in English (*n* = 7).

Materials. The experiment took the form of a Python GUI for the experimental tasks and stimuli, followed by a Google Forms survey. Participants were provided with headphones and volume controls, and the computer was equipped with a Yeti Stereo Microphone for audio recording. The writing prompt participants were asked to respond to (in 3-5 sentences) was the following: "Do you think that a baker should be allowed to refuse to bake a wedding cake for a same-sex couple's wedding when it violates the baker's religious beliefs?" This question was selected to yield responses on both sides within our UC Santa Cruz participant pool, and encourage a sense of lively debate.

The Python GUI contained mid-task survey questions to follow the initial writing task (purportedly while the addressee was finishing their own response). This first mid-task survey included a question on true belief (to select an opposing response). The addressee's response was one of two pre-written responses (each 101 words in length):

> *Response 1.* "No. I don't think that bakers should have the right to deny service based solely on their own religious beliefs. Baking and selling the cake to a gay couple would cause no harm to the baker, and it's not fair for the baker to blatantly discriminate and deny a couple a cake. Even if it's their private business, it would still be discrimination to refuse service on the basis of sexuality. I believe that in the end, it is good for the economy and for the baker to sell to the couple, and the baker is receiving compensation for their product."
>
> *Response 2.* "Yes. I think that bakers should have the right to deny service based on their own religious beliefs. Baking is a creative outlet, and therefore it is not up to the government to decide what kind of art a baker must create or who to sell it to. If it's their private business, they should be able to sell to who they want to, and can reserve the right to refuse service to anyone. I believe that in the end, bakers need to be allowed to maintain their own personal, religious, and political beliefs, to run their business as they choose."

A second mid-task survey included questions to assess attention and belief change. Next, for the rhythm recording task, participants were provided with an egg shaker (A Latin Percussion LP004-GLO), which sat in a bright orange Solo cup on the desk by the computer (for visibility). Participants were led to believe that these first 6 rhythms would be sent to their addressee, when in fact they were only for practice. The first 6 rhythm prompts asked participants to create a rhythm “that sounds like a heartbeat,” “that sounds like a rainstorm,” “that sounds like someone running up stairs,” “that sounds like a lullaby,” “that would fit into a horror movie,” or that “you would dance to,” in randomized order.

The addressee’s recorded rhythms were created ahead of time, and were a combination of real human rhythm recording, and artificial repetition. In this way, the recordings were neither created by a human or a machine, but something in-between. The original audio was recorded with the same equipment and lab booth that participants would be using, and then edited in Audacity such that every track started with 1 second of silence to start. After this, a repeated clip of 2 segments of audio (A and B) alternated ABAB until 10 seconds were up, at which point the audio recording would end abruptly. This was to create the illusion of spontaneity, while the repeated clips were designed to add credence to the notion that a machine learning algorithm might have produced the rhythm. Six such recordings were played two times through without a written prompt description (a total of 20 seconds per recorded stimulus).

The post-experimental survey was a Google Form, designed much the same as the previous two experiments. The first set of questions asked participants about the trust and persuasion of the addressee they interacted with on a Likert scale of 1-7, “Not at all” to “Very much so” (five questions on persuasion, five questions on trust, with one question from each set reversecoded). The next set of questions asked participants to rate their addressee along the humanization scores used in Experiments 1 and 2, on the same scale as the first set of questions (adapted from Haslam, 2006). After that, there was another set of humanization questions that instead asked participants to rate the addressee from 1-7 as “Much less than average” to “Much more than average” (adapted from Schroeder et al., 2017).

Next, participants were asked two questions about their experience with rhythms and egg-shakers, and two questions about experience with and enjoyment of debate. After this, participants were asked a series of questions to assess their attention to detail in understanding task instructions, followed by a reassessment of their belief on the topic of debate (with a long-answer option for elaboration). Participants then completed a self-assessment regarding effort in the task, and were asked a series of optional long-answer questions designed to prompt them to write about any issues or suspicions they may have had about the task.

Design. I used a 2 (“anonymous student participating in a different lab on another UC campus” or “machine-learning algorithm designed to communicate like a UC student”) x 2 (Synchrony “attempt to copy the rhythm as you hear it” or Non-Synchrony “play a response rhythm that is not the same”) between-subjects design. I looked primarily at the participant's humanization ratings of the addressee, trust and persuasion ratings of the addressee, as well belief change on the post-experimental survey. I assessed the participant's experience with rhythm and music, as well as with debate for potential post-hoc analyses.

Procedure. Participants were taken to a single-occupancy computer booth and were told that they would be interacting with either another student or a machine-learning algorithm about a political topic, followed by a rhythm activity. In actuality, the addressee was a collection of pre-recorded responses. The addressee's debate response was designed to be in direct opposition to the participant's response on the first survey questions, following the initial writing task.

The procedure was as follows. First, participants saw the instructions for their particular condition to interact with another student, or to interact with a machine-learning algorithm. Then participants read the debate prompt, followed by a set of survey questions. Next, participants read the presumed addressee's response, and completed another set of survey questions. Finally, participants completed the rhythm task, which consisted of a series of six rhythm instructions for the participant to interpret.

The rationale behind interspersing survey questions within the task was two-fold. The questions both acted as filler tasks to add credibility to the deceit that another agent (human or machine) was taking time to respond to the experimental tasks. The questions also assessed information regarding the participant's true belief (so they could be successfully matched to an opposing statement) as well as other important details. These questions asked whether they were arguing what they truly believed, whether the topic affects people in their life, whether they've debated that topic before, how they feel their addressee responded, whether they disagreed with that response, whether that response made them more likely to want to interact with the addressee further, and reassessment of their own true belief following their addressee's response.

After the first six rhythm prompts, the participant was then asked to either synchronize with their addressee's rhythm (played simultaneously as they recorded), or create a rhythm that was not the same as the one they heard (which did not play as they recorded if they were in the non-synchrony condition). This was repeated 6 times, where the recording they heard was a 10-second recording played two times through (20 seconds total). For the rhythm task, participants could opt to listen to the recorded stimuli or their own recorded audio as many times as they chose, allowing them to re-record as many times as they desired. When ready, they would submit their audio track and move on to the next instruction. Following each synchrony or non-synchrony recording, participants were asked how well they felt they followed the instruction (on a 1-5 scale).

The synchrony and non-synchrony conditions differed based on whether people were moving together with the audio or in sequence with the audio. For the synchrony condition, it was important for the participants to move in synchrony with the audio, because if they were sequential, this behavior could be described as mimicry rather than synchrony. At the same time, for the non-synchrony condition, it was important for the participants to not be producing rhythms together with the audio, because this may lead to people making similar or complementary rhythms as opposed to distinct rhythms.

After the completion of the debate and rhythm task, the participants were asked to fill out a final questionnaire regarding their addressee's humanization characteristics, as well as perceived trust and persuasiveness of the addressee.

Results

The following section presents analyses for an inter-rater reliability test on egg-shaker recordings, a humanization scale comparison, a look at participants who changed their beliefs early in the task, humanization and subscale comparison between groups, as well as confidence, trust, and persuasiveness group comparisons.

Synchrony task instructions. Before conducting the analysis, I asked three independent coders to go through and listen to all of the recordings produced by the participants, when they were asked to either match the rhythm of their addressee, or play a different rhythm. The coders were blind to the condition of each set of audio recordings, and made a decision about whether it seemed the participant was generally trying to match, or not match the original audio recording stimuli. The consensus scores were compared between all three coders, and there was a high level of agreement ($K = .89$, $p < .001$ between raters 1 and 2; $K = .94$, $p < .001$ between raters 2 and 3; $K = .91$, $p < .001$ between raters 3 and 1). Participants who failed to follow instructions were excluded from the following analyses.

Humanization Scale. In order to validate the scale I have been using up to this point (in Experiments 1 and 2), I performed a bivariate correlation between results from my own humanization scale (adapted from Haslam, 2006) and the other version (as reported in Shroeder et al, 2017). For each, I combined all humanization scores into one measure comprising positive and negative qualities for both Human Nature and Human Uniqueness subscales. Both scales resulted in similar overall averages, with my adaptation of the scale performing similarly (*M* = 3.63, *SD* = 0.52) to the other version of the humanization scale (*M* = 3.53, *SD* = 0.45), *r* = .38, *p* < .001. Given that these scales were highly correlated, for the sake of simplicity I discuss results from my own humanization scale for the following analyses.

Unclear or Changed Beliefs. A total of 20 participants expressed that their belief in their own response to the debate prompt was unclear, and that they found their addressee's response convincing, and even changed their mind before participating in the rhythm task. Initially, 35 (37.6%) participants answered "Yes" and 58 (62.4%) responded "No" to the question "Do you think that a baker should be allowed to refuse to bake a wedding cake for a same-sex couple's wedding when it violates the baker's religious beliefs?" Of the 35 who responded "Yes," 11 (31.4%) were unclear on their belief. Fewer of those who responded "No" ended up being unclear on their belief, with just 9 of the initial 58 (15.5%) displaying uncertainty and subsequently changing their response. In order to determine whether to exclude these participants from the analysis, I conducted an independent samples *t*-test for overall humanization scores. I found no significant difference between participants with unclear or changed beliefs (*M* = 3.44, *SD* = 0.55, *95%CI* = [3.21, 3.66], *n* = 20) and participants who had clear and fixed beliefs (*M* = 3.68, *SD* = 0.5, *95%CI* = [3.57, 3.8], *n* = 73), *t*(91) = -1.93, *p* = .06, *95%CI* = [-0.01, 0.51]. For this reason, these participants were included in the following analyses.

Humanization Main Effects. For combined humanization scores, main effects were observed for both backstory, but not for synchrony. Participants who thought they were interacting with a human (another student), they rated their addressee more highly on combined humanization (*M* = 3.73, *SD* = 0.51, *95%CI* = [3.59, 3.88]) than participants who thought they were interacting with a computer (machine learning algorithm), (*M* = 3.52, *SD* = 0.51, *95%CI* = [3.36, 3.67]), *F*(1, 89) = 4.27, *p* = .04. While non-significant, participants who engaged in synchrony also gave slightly higher combined humanization ratings to their addressee (*M* = 3.71, *SD* = 0.51, *95%CI* = [3.57, 3.86]) than participants who did not, (*M* = 3.54, *SD* = 0.51, *95%CI* = [3.39, 3.69]), *F*(1, 89) = 2.74, *p* = .1.

Human Nature and Human Uniqueness. To better understand the group differences in humanization, I examined the effect of backstory and synchrony on positive and negative Human Nature and Human Uniqueness ratings of the addressee. Of these, a significant main effect was found for negative Human Nature traits, and a significant interaction effect was found for positive Human Uniqueness traits. For negative ratings of Human Nature, when participants were told they were interacting with another student, they rated their addressee more highly for negative traits (*M* = 3.26, *SD* = 3.26, *95%CI* = [2.98, 3.55]) than when they were told they were interacting with an algorithm, (*M* = 2.7, *SD* = 2.71, *95%CI* = [2.41, 3]), *F*(1, 89) = 7.42, *p* = .008.

As for interaction effects, when participants were told they were interacting with another student, they rated their addressee more highly on positive Human Uniqueness qualities when they synchronized in the rhythm activity (*M* = 4.22, *SD* = 0.79, *95%CI* = [3.89, 4.55]) than when they were asked to make a different rhythm (*M* = 3.78, *SD* = 0.87, *95%CI* = [3.44, 4.13]). Conversely, when participants were told they were interacting with an algorithm, they tended to rate their addressee less highly on positive Human Uniqueness qualities when they synchronized in the rhythm activity (*M* = 4.01, *SD* = 0.85, *95%CI* = [3.67, 4.35]) than when they were asked to make a different rhythm (*M* = 4.32, *SD* = 0.82, *95%CI* = [3.97, 4.67]), *F*(1, 89) = 4.75, *p* = .03.

Confidence. In a comparison of self-reported confidence, following each rhythm recording, Participants in the synchrony group reported lower confidence in their ability to follow instructions (*M* = 3.93, *SD* = 0.53, *95%CI* = [3.78, 4.08]) than those in the non-synchrony group (*M* = 4.57, *SD* = 0.53, *95%CI* = [4.41, 4.73]), *F*(1, 89) = 33.7, *p* < .001. There were no other significant differences between groups.

Trust of Addressee. While I expected to observe higher ratings of trust for participants who were told their partner was human, and for participants who engaged in a synchrony task, no significant group differences were observed, *F*(1, 89) = 4.27, *p* = .04.

Persuasiveness of Addressee. Although there were no observed differences in self-reported trust of the addressee, there was a significant interaction effect for persuasiveness. When participants were told they were interacting with another student, they rated their addressee more highly for persuasiveness when they synchronized in the rhythm activity (*M* = 3.27, *SD* = 0.95, *95%CI* = [2.84, 3.7]) than when they were asked to make a different rhythm (*M* = 2.57, *SD* = 1.36, *95%CI* = [2.13, 3.02]). Conversely, when participants were told they were interacting with an algorithm, they rated their addressee less highly for persuasiveness when they synchronized in the rhythm activity (*M* = 2.94, *SD* = 1.05, *95%CI* = [2.49, 3.39]) than when they were asked to make a different rhythm (*M* = 3.63, *SD* = 0.90, *95%CI* = [3.17, 4.09]), *F*(1, 89) = 9.56, *p* = .003.

Discussion

Experiment 3 included the same humanization measure as in previous experiments, but also compared another version of this scale, implications of belief change, group differences in humanization of the anonymous agent, confidence in synchrony, and finally the trust and persuasion findings.

Humanization Scale. There were various reasons I opted to use my own adaptation of Haslam's (2006) humanization scale. One reason was for consistency, in order to do comparisons across Experiments 1, 2, and 3 more easily. Another reason was the one stated in the results: namely, that the two scales were highly correlated and did not seem to differ in central tendency.

The final reason has more to do with the context of the research, in an human-computer interactions study. In the other adaptation of the scale (Schroeder et al., 2017), the questions are set up to ask people to compare their addressee to others in a way that doesn't make a lot of sense in a context where you are judging the behavior of a machine. One participant, who happened to be in the human backstory and synchrony group, said: “I didn't really understand the questions where the answers were ‘more than average, less than average’ so I might have answered in a way that isn't reflective of what I think because I was kind of confused by the answer choices in relation to the question. I'm not sure if the answers I chose represent what I meant to say or choose.” Because there seemed to be confusion with this style of question, and because the scales seemed to be correlated strongly enough, I opted to focus my analysis on the version I had been using previously.

Unclear or Changed Beliefs. Interestingly, approximately 21% of participants changed their minds after reading their partner’s response to the debate question, even before engaging in the synchrony activity. Because it would appear that these participants who changed their minds after the initial exposure to the opposing belief did not significantly differ in their overall humanization rating of the addressee, they were not excluded from the analysis.

Looking at the directions in changed beliefs, more students (62.4%) initially responded with a “No” to the forced-choice question “Do you think that a baker should be allowed to refuse to bake a wedding cake for a same-sex couple’s wedding when it violates the baker’s religious beliefs?” These same students tended to also be more resolute in their beliefs (with 84.5% not changing their minds). In the post-experimental survey, when I asked what students thought the study was about, one replied “I guess the connection between following a rhythm and homophobia?” It would appear that this pattern of responses may have more to do with the political climate at UCSC than anything else, as one belief was more popular.

In my initial design, I wanted to make sure that it would be believable to receive an opposing response, and not make the participant doubt the validity of their partner's existence. Although it would have been ideal to have a 50-50 split, no students expressed disbelief about their addressee's existence on the basis of making an unusual argument. By setting participants up with an opposing viewpoint to read, I was hoping to set the groundwork for initial dehumanization of the addressee, in order to establish a clear intervention strategy through the use of backstories and synchrony. Still, it is perhaps a testament to the utility of these strategies that even participants who initially agree can begin to humanize their addressee to a greater extent.

It also goes to show that beliefs are not fixed, and even interactions with anonymous strangers (human or machine) can lead to belief change. That said, if this methodology is to be improved upon, I would recommend the use of multiple debate prompts, followed by a survey designed to assess which beliefs are strongest, in order to match each participant up with a debate response with which they are less inclined to agree.

Humanization and Confidence. Even while interacting with an anonymous addressee, just being told that it was another UC student (human) or a machine learning algorithm (computer) was enough to affect overall humanization scores. In reading survey responses, it's apparent that participants began to draw inferences about their addressee, based on backstory alone.

Participants in the human backstory group, for example, said the following things when making inferences about their addressee: “Their knowledgable *[sic]* in the subject, and passionate about their beliefs. They seem to be creative based on their rhythm making” and “I can infer that he/she is creative and has a clear understanding of politics.”

On the other hand, participants in the computer backstory group said the following: “It felt a little computerized” and “They seem to use logic in their arguments, instead of emotions.” It is worth re-emphasizing that the stimuli produced by their so-called addressee were the same across conditions (the only difference being the pro/con argument, which did not vary systematically by condition).

Participants were clearly more confident about their ability to follow instructions in the non-synchrony condition, which might be part of the reason that synchrony effects, while trending, were not significant. One participant in the computer backstory and synchrony condition said, “it was hard to copy its rhythm exactly.” Synchrony may have boosted the perception of an anonymous addressee’s humanlike qualities, but perhaps in a more stressful way than was intended. It is possible that the stress of participating in a synchrony task that required too much effort may have reduced any possible benefits that a rhythm-based activity might otherwise offer.

There was one main effect for a humanization subscale. For negative Human Nature traits (rating the addressee as jealous, nervous, impatient, distractible, and aggressive), the supposedly human agent received higher scores than the computer agent. Between Human Nature and Human Uniqueness, the former is related more to animalistic characteristics, and the latter is related to more mechanistic characteristics. Given that humans are more like an animal than computers are, it makes a certain amount of sense that Human Nature scores, even for negative traits, would be higher for human rather than machine agents.

There was also an interaction effect for positive Human Uniqueness traits (rating the addressee as humble, thorough, organized, polite, and broadminded). Participants who synchronized with a supposedly human agent gave higher ratings for these traits than those who didn't synchronize, and participants who synchronized with a supposedly computer agent tended to give lower ratings for these traits than those who didn't synchronize. The reason for this interaction effect is unclear. It is possible that participants appreciate certain mechanistic qualities of a human they are trying to synchronize with (in a rather challenging rhythm-matching task), but also value a computer agent to a greater extent when they are *not* asked to copy what they may interpret as boring and mechanistic actions, or in some way made to compare themselves to a supposedly machine agent.

Trust and Persuasiveness of Addressee. For ratings of trust, it is possible that participants were less keen to respond to questions in the survey that might cast their addressee in a disparaging light (possibly even for the computer agent). In some of the written survey responses, participants who interacted with a supposedly human agent offered reserved responses such as: “The addressee is a person who has an opinion and shared a response” and “I did feel a little self conscious answering the survey, like I was worried about what my partner was saying about me." Meanwhile, participants who interacted with a supposedly machine agent responded in a similar manner, with comments such as: "They are an AI that can use collected information to seem more human" and "it has learned an array of different rhythms and cadences."

As for persuasiveness, participants gave higher ratings to humans they synchronized with, and machines they did not synchronize with. As discussed previously, the reason for this is unclear. Synchrony and humanness were expected to lead to higher persuasiveness ratings, but the fact that non-synchrony in computer agents would lead to higher persuasiveness ratings than synchrony in computer agents is puzzling. Again, it is possible that people simply do not enjoy being asked to perform the same job as a machine. Even beyond being unsettled by machines that are supposedly taking human jobs, being saddled with a machine's job can lead to a certain degree of resentment. If a person is asked to synchronize with a machine, they may develop a dislike for that agent, and therefore feel less willing to change their initial beliefs in order to find common ground with that machine. It is also possible that robots that are viewed more mechanistically (recall that Human Uniqueness followed a similar pattern to persuasiveness ratings for the four groups) are also potentially viewed as more impartial, and therefore may have more persuasive arguments (even if the agents themselves are no more trustworthy)

In the next, and final analysis, I will do a cross-experimental comparison of humanization scores. This comparison is exploratory in nature, and is provided for the benefit of the reader. It is quite likely that further development of a measure is warranted, for use in the development of likeable machine agents in various social contexts.

Cross-Experimental Comparison of Humanization Scores

In this section, I outline a cross-experimental comparison of all conditions from Experiments 1, 2, and 3. This analysis made use of pre-existing data from all experiments up to this point. There was no overarching hypotheses, as this analysis was primarily conducted for exploratory purposes, in hopes of informing the use of scales similar to that of Haslam's (2006) humanization measure in the context of human-computer and human-robot interaction.

Results

Significant group differences were observed between experimental conditions, $F(9, 344) = 23.06$, $p < .001$ (Figure 4). Of particular note, all groups from Experiment 1 rated the depictions of robots significantly more highly on humanization traits than both robot groups from Experiment 2. Additionally, they also rated the depictions of robots more highly on humanization traits than both machine agent groups from Experiment 3. Both TheraBot groups also rated machine agents more highly than both human agent groups from Experiment 3. The *RoboShrink before mimicry* group also rated the machine agent more highly than the *human non-synchrony* group in Experiment 3, while the *TheraBot after mimicry* group rated the machine agent more highly than the *RoboShrink before mimicry* group. The human-synchrony group from Experiment 3 received higher humanization scores, however, than the *non-synchronous robot* condition from Experiment 2.

Figure 4.

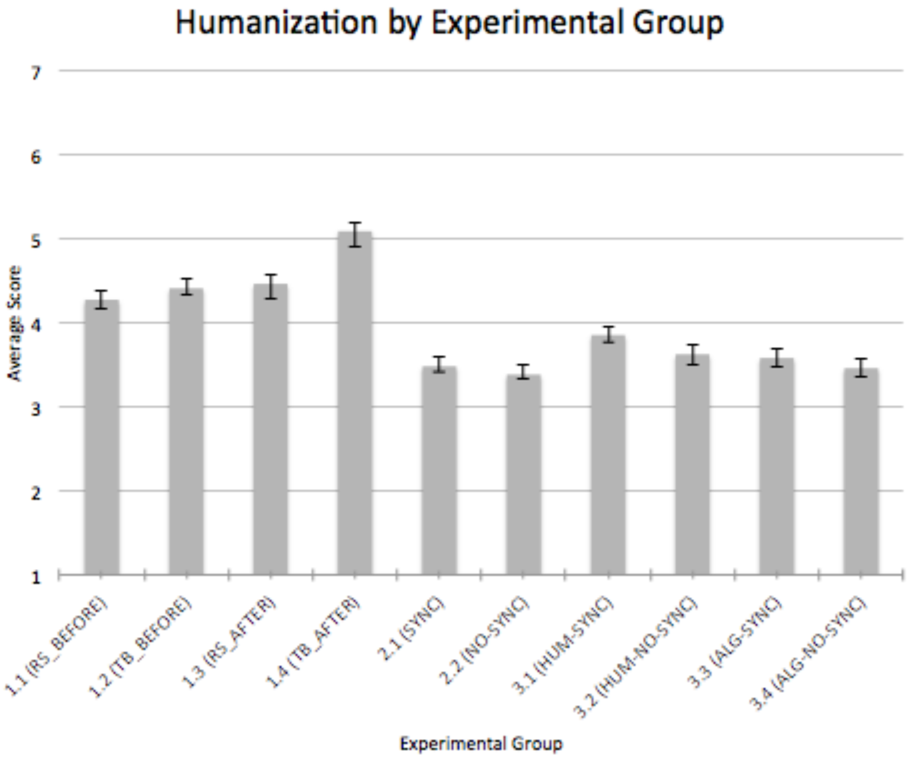


Figure 4. Cross-experimental comparison of humanization scores, from left to right: 1.1 (RoboShrink before mimicry), 1.2 (TheraBot before mimicry), 1.3 (RoboShrink after mimicry), 1.4 (TherabBot after mimicry), 2.1 (Synchrony with Mabu), 2.2 (Non-synchrony with Mabu), 3.1 (Human and synchrony), 3.2 (Human and non-synchrony), 3.3 (Algorithm and synchrony), 3.4 (Algorithm and non-synchrony).

Discussion

While this cross-experimental comparison of humanization scores is exploratory, it is still interesting to consider what it can tell us about the humanization measure itself. Overall, Experiment 1 had the highest ratings of humanization of all 3 experiments (consistently higher than the actual, physical robots in Experiment 2, and higher than both machine groups from Experiment 3).

General Discussion

When it comes to interactions with other humans, perceiving another person as different from oneself or the group can sometimes result in the perception that an individual is somehow less than human (Haslam, 2006; Haslam & Loughnan, 2014). In some cases, this can result in something as extreme as animalistic or mechanistic dehumanization, when fundamental traits of human uniqueness or nature, respectively, are denied (Crawford, Modri, & Motyl, 2013). There is converging evidence that similarity attraction is also at play in human-computer interactions, with people preferring machine agents who display characteristics that are more human or more well-matched to their own personality characteristics (Nass & Brave, 2005). As computer-mediated communication becomes increasingly common, and the boundaries between human and machine interactions become increasingly unclear, it is important to assess in which ways these humanization measurements may also apply to non-human agents.

Other research suggests that there are some possible motion synchrony activities that can induce a sense of kinship and trust between strangers (Wiltermuth & Heath, 2009). While some research has implied that the same may be true of human-robot interactions, this research has been observational and not entirely generalizable (Michalowski, Sabanovic, & Kozima, 2007; Michalowski, Simmons, & Kozima, 2009). My research works to bridge this gap in understanding, in terms of the humanizing impacts of motion synchrony in human-computer and human-robot interactions.

Experiment Review

In three experiments involving human-computer interactions with non-human virtual agents, I investigated the impact of synchrony exercises on humanization trait ratings. Each experiment had a focus on some element of health. Experiment 1 involved training non-human agents to correctly assess mental health, while Experiment 2 involved the use of a robot motivational coaches for physical health, and Experiment 3 had participants debating human or non-human agents online in the context of healthy interpersonal behaviors over the internet. The three experiments were designed to be successively more true to real-life human-computer interactions, with Experiment 1 using static robot depictions, Experiment 2 using functional robots that can move, and Experiment 3 simulating an online, text-based conversation with an anonymous agent (either human or machine).

In this research, I was seeking to test three main hypotheses, related to humanization, synchrony, and backstories: (1) that a humanization measure (adapted from Haslam, 2006) can be used in human-computer and human-robot interaction studies; (2) that behavioral matching, or synchrony, can lead to higher humanization ratings for machine agents, as well as human agents; (3) that while humans will receive higher humanization ratings than robots that perform the same actions, collaborative synchrony will boost humanization ratings for all agents.

Experiment 1 provides novel evidence that humanization scales may be used as a measure of impression change after a behavioral matching task intervention (mimicry training). Humanization ratings were found to correspond to preference for one robot over another, and participating in the optional mimicry task resulted in higher humanization scores for TheraBot (the helper bot), but not for RoboShrink (the replacement bot). Results from Experiment 1 indicate that a mimicry task can lead to improvements in the humanization ratings of a robot that is presented as a helper rather than replacement robot. Both RoboShrink and TheraBot received higher overall scores for Human Uniqueness traits than Human Nature traits (both positive and negative items), which corresponds to more mechanistic than animalistic impressions (rightly so, given that both were presented as robots).

In Experiment 2, participants overwhelmingly preferred to respond positively to the robot. When given the option, they usually selected one of the two positive and affirmative responses (Positive–Warm or Positive–Cold). The synchrony motion activity did have an impact in how participants chose to respond to the robot directly, with warmer responses during and following the synchrony activity than without a synchrony activity. People who participated in the motion synchrony task and performed the stretching exercises while the robot moved along with them tended to respond more warmly to the robot following the activity than participants who only heard the robot giving them directions on which stretches to perform. Humanization ratings and memory did not vary across conditions. People rated robots they moved in- and out-of-synch with similarly. They also remembered information about the robots similarly.

In Experiment 3, participants were led to believe they were interacting with a human or machine agent who disagreed with them on a topic of political debate. Next, they were asked to perform a synchronous or non-synchronous motion activity with an egg-shaker. Anonymous agents with a human backstory received higher humanization scores overall. Interestingly, while synchrony led to higher ratings for humanization for human agents, participants gave higher ratings of positive Human Uniqueness to machines that they did not synchronize with. The human synchrony condition also led to higher perceived persuasiveness in an agent than human non-synchrony, but machine synchrony led to lower perceived persuasiveness in an agent than non-synchrony. Taken together, it seems clear that backstories and synchrony interventions can affect not only the perceived humanness of an agent, but also the persuasiveness of that agent. There is also some indication that machine agents would be best advised to encourage creative interactions, rather than repetitive copy-cat interactions, in order to be considered more persuasive and humanlike (if that is the goal).

Cross-Experiment Humanization Comparison

In a cross-experimental comparison of all the groups from each of these three experiments, there appeared to be certain differences between how student participants rated the agent they interacted with on overall humanization. For example, the illustrated depictions of robots in Experiment 1 (with the exception of RoboShrink following the mimicry task) all received higher humanization scores than the two robot conditions from Experiment 2. Both TheraBot conditions also outperformed all but the human-synchrony agent from Experiment 3 (which in turn received higher humanization than the non-synchronous robot in Experiment 2).

It is hard to say for certain, but it could be that students studying Psychology may have been particularly drawn to the idea of working to train robots that were designed to aid in providing therapy to those who need it. The mental health aspect of the robot's purpose may have resonated with them to such an extent that they began to humanize the robots, almost as in-group members.

Looking at existing research, there is reason to believe that, in some cases, more impoverished machine agents can lead to more positive judgements (Lee, 2010; Nowak & Biocca, 2004). This could in part have something to do with the uncanny valley hypothesis, which suggests that machine agents can benefit from similarity to humans, up to the point at which they become unsettling (Mori, MacDorman, & Kageki, 2012). The fact that all of the experimental groups in Experiment 1 gave higher humanization scores to the drawings of robots – compared to not only the robots of Experiment 2, but also the machine agents of Experiment 3 – implies that perhaps the impoverished image is a safer bet for human-computer interactions.

It could also be that people are more accustomed to interfacing with cartoons and illustrated depictions of robots in the media. This familiarity could in turn lead to a sense of comfort and work to boost the perceived humanization of the robot agents. Whereas in Experiment 2, while the robots were quite popular with participants, the presence of physical robots can still be seen as threatening to some. As for Experiment 3, participants were interacting with an anonymous other. The drawings of cute robots could simply pose less of a threat.

Conclusions

The driving question in this research was whether people can develop more positive impressions (higher humanization scores) for nonhuman and human agents alike when they are asked to engage in a motion synchrony task with these agents. This research lends support to the use of the collaborative synchrony in fostering positive intergroup relations (for example, in visual communication such as sign language and gestures, and musical or rhythm-based games). While it is as yet unclear exactly what the effects of motion synchrony are on the perception of machine agents, these experiments reveal that non-threatening machine agents can benefit from engaging users in motion-based interaction tasks, both in terms of humanization score improvements (as seen in Experiment 1) and in terms of warmer user responses (as seen in Experiment 2).

When it comes to agents that are potentially seen as outgroup members, synchrony can lead to improvements in humanization ratings for machine and human agents alike, but machine agents are more persuasive when users are not asked to engage with them in a rhythm-based motion synchrony task (as seen in Experiment 3). One possible interpretation is that people place greater credence in the behaviors of a machine when they are not asked to engage with it repetitively, but creatively. Another interpretation is that people place higher confidence in machines that appear distant, and therefore impartial. Feeling less of a bond with the machine (that is, not participating in a synchrony task), could lead to a sense that the machine is more logical. For example, in Experiment 1, RoboShrink more often misled participants who favored that machine design. It is possible that RoboShrink was more persuasive precisely because it was seen as machine-like, logical, and impartial.

Also in Experiment 1, participants that opted to train the first robot they were introduced to tended to rate TheraBot more highly after the mimicry task, while the same was not true for those that trained RoboShrink. It is possible that emotionally distant and threatening machine agents don't benefit from a synchrony activity to the same extent as endearing robots that do not pose a threat. The machine agent in Experiment 3, set up initially to disagree with the participant, could still be seen as enough of a threat that a participant becomes hyper-aware of the machine's negative qualities when asked to synchronize with it. Regardless of interpretation, it is clear that synchrony has a different impact on the perceived persuasiveness of machines than humans.

In conclusion, these experiments revealed the following: (1) There is some information that can be captured about a participant's feelings towards a machine or supposedly human agent, through the use of a humanization measure (adapted from Haslam, 2006). It is important to bear in mind that even the negative traits on the scale are important aspects of what it is to be human, and should not be dis-included in the final analysis. (2) Motion synchrony can indeed contribute to higher humanization ratings for supposedly human, as well as machine agents. However, in certain circumstances, when interacting with a physical robot (as in

Experiment 2), synchrony interventions seem to have less of an effect on perceived humanness of that agent. (3) The perceived humanness of an agent (backstory) does impact their humanization scores, but collaborative synchrony can also boost these scores, mostly for human agents (as in Experiment 3).

Through this research, I sought to better understand how to encourage positive interactions between groups that have effectively dehumanized one another, perhaps due to a fear that the other group poses a threat to their well-being (For example, a threat to their job security). Studying the interactions of humans and machines can have a beneficial impact on human-human interaction too: we can come to a better understanding of how to interact with those who are dehumanized, or seen as belonging to an outgroup. If synchrony-boosting exercises can help break down self-other distinctions, perhaps this will result in more positive interactions and less stigma between people.

References

Aharoni, E., Fridlund, A.J. (2007). Social reactions toward people vs. computers: How mere labels shape interactions. *Computers in Human Behavior, 23,* 2175-2189. DOI: 10.1016/j.chb.2006.02.019.

Alibali, M.W., Heath, D.C., Myers, H.J. (2001). Effects of visibility between speaker and listener on gesture production: Some gestures are meant to be seen. *Journal of Memory and Language, 44,* 169-188. DOI: 10.1006/jmla.2000.2752.

Bar, M., Neta, M., Linz, H. (2006). Very first impressions. *Emotion, 6*(2), 269-278. DOI 10.1037/1528-3542.6.2.269.

Bavelas, J., Gerwing, J., Sutton, C., Prevost, D. (2007). Gesturing on the telephone: Independent effects of dialogue and visibility. *Journal of Memory and Language, 58*, 495-520. DOI: 10.1016/j.jml.2007.02.004.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction (TOCHI), 12*(2), 293-327.

Branigan, H.P., Pickering, M.J., Pearson, J., Mclean, J.F. (2010). Linguistic alignment between people and computers. *Journal of Pragmatics, 42*(9), 2355-2368. DOI: 10.1016/j.pragma.2009.12.012.

Buck, D.M., Plant, A. (2010). Interorientation interactions and impressions: Does the timing of disclosure of sexual orientation matter? *Journal of Experimental Social Psychology, 47,* 333-342. DOI 10.1016/j.jesp.2010.10.016.

Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.

Cooper, T. (2011, August). *The Case Study of Susie: Bipolar/ Disorder.* Retrieved from http://criminologyjust.blogspot.com/2011/08/case-study-of-susie-bipolar-i-disorder.html#.WgEI%205RNSzVo.

Crawford, J.T., Modri, S.A., Motyl, M. (2013). Bleeding-heart liberals and hard-hearted conservatives: Subtle political dehumanization through differential attributions of human nature and human uniqueness traits. *Journal of Social and Political Psychology, 1*(1), 86-104. DOI 10.5964/jspp.v1i1.184.

de Melo, C.M., Gratch, J., Carnevale, P.J. (2014). Humans vs. computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing,* 1-11.

Fitzpatrick, K.K., Darcy, A., Vierhile, M. (2017). Delivering Cognitive Behavior Therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (Woebot): A randomized controlled trial. *JMIR Mental Health, 4*(2). doi: 10.2196/mental.7785

Galvão Gomes da Silva, J., Belpaeme, T., Kavanagh, D. J., Taylor, L., Beeson, K., & Andrade, J. (2018). Experiences of a motivational interview delivered by a robot: A qualitative study. *Journal of medical Internet research, 20*(5), e116.

Giles, H., Coupland, N., Coupland, J. (1991). Accommodation theory: Communication, context, and consequence. *Contexts of Accomodation: Developments in Applied Sociolinguistics, 1-68.*

Gillig, P. M. (2009). Dissociative Identity Disorder: A Controversial Diagnosis. *Psychiatry (Edgmont),* 6 (3), 24–29. Retrieved from ncbi.nlm.nih.gov

Gray, K., & Wegner, D.M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition*, *125*(1), 125-130.

Hammond, A.A., D'Arcey, J.T., Larson, A.S. & Fox Tree, J.E. (2017). The more sarcastic you are, the more sarcastic you think others are. Poster presented at the Psychonomic Society's 58th Annual Meeting. Vancouver, British Columbia, Canada.

Hammond, A.A., D'Arcey, J.T., Larson, A.S., Fox Tree, J.E. (in prep). The Sarchasm: Sarcasm Production and Identification in Spontaneous Conversation.

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10*(3). 252-264.

Haslam, N., & Loughnan, S. (2014). Dehumanization and Infrahumanization. Annual Review of Psychology 65, 399-423.

Isbister, K. (2012). How to Stop Being a Buzzkill: Designing Yamove!, A Mobile Tech Mash-up to Truly Augment Social Play. Keynote presentation, abstract included in *Proceedings of MobileHCI 2012*, San Francisco, CA.

Isbister, K., Nass, C. (2000). Consistency of personality in interactive characters: Verbal cues, non-verbal cues, and user characteristics. *Int. J. Human-Computer Studies, 53,* 251-267. DOI: 10.1006/ijhcs.2000.0368.

Ito, T.A., Urland, G.R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology, 85*(4), 616-626. DOI 10.1037/0022-3514.85.4.616.

Iverson, J.M., Goldin-Meadow, S. (2001). The resilience of gesture in talk: Gesture in blind speakers and listeners. *Developmental Science, 4*(4), 416-422.

Krumhuber, E., Manstead, A.S.R., Cosker, D., Marshall, D., Rosin, P.L. (2008). Effects of dynamic attributes of smiles in human and synthetic faces: A simulated job interview setting. *Journal of Nonverbal Behavior, 33*, 1-15. DOI 10.1007/s10919-008-0056-8.

Kuriki, S., Tamura, Y., Igarashi, M., Kato, N., Tamami, N. (2016). Similar impressions of humanness for human and artificial singing voices in autism spectrum disorders. *Cognition*, *153*, 1-5.

Lane, C. (November 02, 2017). *Case Studies*. Retrieved from http://www.psyweb.com/Casestudies/CaseStudies.jsp#3

Larson, A.S., Fox Tree, J.E. (in review). Framing, more than speech, affects how machine agents are perceived.

Lassiter, G.D., Irvine, A.A. (1986). Videotaped confessions: The impact of camera point of view on judgments of coercion. *Journal of Applied Social Psychology, 16*(3), 268-276.

Lassiter, G.D., Beers, M.J., Geers, A.L., Handley, I.M., Munhall, P.J., Weiland, P.E. (2002). Further evidence of a robust point-of-view bias in videotaped confessions. *Current Psychology, 21*(3), 265-288.

Lassiter, G.D., Geers, A.L., Handley, I.M., Weiland, P.E., Munhall, P.J. (2002). Videotaped interrogations and confessions: A simple change in camera perspective alters verdicts in simulated trials. *Journal of Applied Psychology, 87*(5), 867-874. DOI 10.1037//0021-9010.87.5.867.

Lee, K.M., Nass, C.. (2004). The multiple source effect and synthesized speech. *Human Communication Research, 30*(2), 182-207.

Lee, E-J. (2010). The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior*, *26*(4), 665-672.

Louwerse, M.M., Dale, R., Bard, E.G., Jeuniaux, P. (2012). Behavior matching in multimodal communication is synchronized. *Cognitive Science, 36*, 1404-1426. DOI: 10.1111/j.1551-6709.2012.01269.x.

Macrae, C. N., Duffy, O. K., Miles, L. K., & Lawrence, J. (2008). A case of hand waving: Action synchrony and person perception. *Cognition, 109*(1), 152–156.

Manning, J. S. (1999). Valproate in Bipolar Disorder: Case Examples From Family Practice. *The Primary Care Companion to The Journal of Clinical Psychiatry,* 01 (03), 71-73. doi:10.4088/pcc.v01n0303.

Michalowski, M. P., Sabanovic, S., & Kozima, H. (2007, March). A dancing robot for rhythmic social interaction. In *Human-Robot Interaction (HRI)*

2nd ACM/IEEE International Conference on 2007 March 9 (pp. 89-96). IEEE.

Michalowski, M. P., Simmons, R., & Kozima, H. (2009, September). Rhythmic attention in child-robot dance play. In *Robot and Human Interactive Communication RO-MAN, The 18th IEEE International Symposium on 2009 Sep 27* (pp. 816-821). IEEE.

Moon, Y., Nass, C. (1996). How “real” are computer personalities? Psychological responses to personality types in human-computer interaction. *Communication research*, *23*(6), 651-674.

Mori, M., MacDorman, K.F., Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine, 19*(2), 98-100.

Morkes, J., Kernal, H.K., Nass, C. (1998). Humor in task-oriented computer-mediated communication and human-computer interaction. *CHI 98, 18*(23), 215-216.

Motley, M.T., Camden, C.T. (1988). Facial expression of emotion: A comparison of posed expressions versus spontaneous expressions in an interpersonal communication setting. *Western Journal of Speech Communication, 52*(1), 1-22. DOI: 10.1080/10570318809389622.

Nagels, A., Kircher, T., Steines, M., Straube, B. (2015). Feeling addressed! The role of body orientation and co-speech gesture in social communication. *Human Brain Mapping, 36*, 1925-1936. DOI: 10.1002/hbm.22746.

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

Nass, C. Lee, K.M. (2000). Does computer-generated speech manifest personality? An experimental test of similarity-attraction. *CHI Letters, 2*(1), 329-336. DOI: 10.1145/332040.332452.

Nass, C.I., Lombard, M., Henriksen, L., Steuer, J. (1995). Anthropocentrism and computers. *Behavior and Information Technology, 14*(4), 229-238.

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

Nass, C., Moon, Y. & Carney, P. (1999). Are people polite to computers? Responses to computer-based interviewing systems. *Journal of Applied Social Psychology, 29*(5). 1093–1110.

Nass, C., Moon, Y., Fogg, B.J., Reeves, B., Dryer, D.C. (1995). Can computer personalities be human personalities? *Int. J. Human-Computer Studies, 43,* 223-239.

Neff, M., Wang, Y., Abbott, R., Walker, M. (2010). Evaluating the effect of gesture and language on personality perception in conversational agents. *Intelligent Virtual Agents*, 222-235.

Nowak, K.L., Biocca, F. (2004). The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence, 12*(5), 481-494.

Olivola, C.Y., Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior, 34*, 83-110. DOI 10.1007/s10919-009-0082-1.

Pacilli M.G., Roccato, M., Pagliaro, S., Russo, S. (2015). From political opponents to enemies? The role of perceived moral distance in the animalistic dehumanization of the political outgroup. *Group Processes & Intergroup Relations,* 1-14. DOI: 10.1177/1368430215590490.

Pettigrew, T.F., Tropp, L.R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology, 90*(5), 751-783. DOI 10.1037/0022-3514.90.5.751.

Rockwell, P. (2000). Lower, slower, louder: Vocal cues of sarcasm. *Journal of Psycholinguistic Research, 29*(5), 483-495.

Rolls, G. (2015). The Boy Who Couldn't Stop Washing: a story of OCD. In Taylor & Francis (Ed.). *Classic Case Studies in Psychology*. (241-246). New York, NY.

Schectman, N., Horowitz, L.M. (2003). Media inequality in conversation: How people behave differently when interacting with computers and people. *In CHI 2003: New Horizons, 5*(1), 281-288.

Schroeder, J., Kardas, M., & Epley, N. (2017). The Humanizing Voice: Speech reveals, and text conceals, a more thoughtful mind in the midst of disagreement. *Psychological Science, 28*(12), 1745-1762.

Shah, N., & Nakamura, Y. (2010). Case Report: Schizophrenia Discovered during the Patient Interview in a Man with Shoulder Pain Referred for Physical Therapy. *Physiotherapy Canada*, 62 (4), 308-315. http://doi.org/10.3138/physio.62.4.308.

Sebanz, N., Bekkering, H., & Knoblich, G. (2006). Joint action: bodies and minds moving together. *Trends in Cognitive Sciences, 10*(2), 70–76.

Seo, S., Geiskkovitch, D., Nakane, M., King, C., & Young, J. (2015). Poor Thing! Would you feel sorry for a simulated robot? A comparison of empathy toward a physical and simulated robot. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*.

Sirkin, D., Mok, B., Yang, S., & Ju, W. (2015, March). Mechanical ottoman: how robotic furniture offers and withdraws support. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction* (pp. 11-18). ACM.

Spitzer, R. L., Skodol, A. E., & Gibbon, M. (2002). *DSM-IV-TR Casebook: A Learning Companion to the Diagnostic and Statistical Manual of Mental Disorders*. Washington, D.C.: American Psychiatric Publishing.

Taggart, W., Turkle, S., Kidd, C.D. (2005). An interactive robot in a nursing home: Preliminary remarks. In *Towards social mechanisms of android science: a COGSCI workshop*.

Takayama, L., Ju, W., & Nass, C. (2008, March). Beyond dirty, dangerous and dull: what everyday people think robots should do. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction* (pp. 25-32). ACM.

Tetlock, P.E. (1983). Accountability and the perseverance of first impressions. *Social Psychology Quarterly, 46*(4), 285-292.

Todorov, A. & Porter, J.M. (2014). Misleading first impressions: Different for different facial images of the same person. *Psychological Science, 25*(7), 1404-1417. DOI 10.1177/0956797614532474.

Tolins, J., Liu, K., Neff, M., Walker, M., & Fox Tree, J. E. (2016). A verbal and gestural corpus of story retellings to an expressive embodied virtual character. *Proceedings of the International Conference on Language Resources and Evaluation*, Portorož, Slovenia, pp. 3461-3468.

Torrey, C., Fussell, S.R., Kiesler, S. (2013). How a robot should give advice. *In HRI 2013 Proceedings,* 275-282.

Torrey, C., Fussell, S.R., Kiesler, S. (2013). How a robot should give advice. *In HRI 2013 Proceedings,* 275-282.

Trull, T. J., & Prinstein, M. J. (2013). *Clinical Psychology*. Singapore: Cengage Learning Asia Pte Ltd.

Unnever, J.D., Cullen, F.T. (2009). Empathetic identification and punitiveness: A middle-range theory of individual differences. *Theoretical Criminology, 13*, 283. DOI 10.1177/1362480609336495.

Walker, E.J., Risko, E.F., Kingstone, A. (2014). Fillers as signals: Evidence from a question-answering paradigm. *Discourse Processes, 51*(3), 264-286. DOI: 10.1080/0163853X.2013.862478.

Walters, M.L., Syrdal, D.S., Koay, K.L., Dautenhahn, K., te Boekhorst, R. (2008). Human approach distances to a mechanical-looking robot with different robot voice styles. *Proceedings to the 17th IEEE International Symposium* *on Robot and Human Interactive Communication, Technishe Universität München, Munich, Germany.* DOI: 10.1109/ROMAN.2008.4600750.

Wiltermuth, S.S., Heath, C. (2009). Synchrony and cooperation. *Psychological Science, 20*(1), 1-5.

Zuckerman, M., Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior, 17*(2), 119-135.