Journal of Experimental Psychology: General

Mistaking Minds and Machines: How Speech Affects Dehumanization and Anthropomorphism

Juliana Schroeder and Nicholas Epley

Online First Publication, August 11, 2016. http://dx.doi.org/10.1037/xge0000214

Journal of Experimental Psychology: General

0096-3445/16/$12.00 http://dx.doi.org/10.1037/xge0000214

Mistaking Minds and Machines: How Speech Affects Dehumanization and Anthropomorphism

Juliana Schroeder
University of California, Berkeley

Nicholas Epley
University of Chicago

Treating a human mind like a machine is an essential component of dehumanization, whereas attributing a humanlike mind to a machine is an essential component of anthropomorphism. Here we tested how a cue closely connected to a person's actual mental experience—a humanlike voice—affects the likelihood of mistaking a person for a machine, or a machine for a person. We predicted that paralinguistic cues in speech are particularly likely to convey the presence of a humanlike mind, such that removing voice from communication (leaving only text) would increase the likelihood of mistaking the text's creator for a machine. Conversely, adding voice to a computer-generated script (resulting in speech) would increase the likelihood of mistaking the text's creator for a human. Four experiments confirmed these hypotheses, demonstrating that people are more likely to infer a human (vs. computer) creator when they hear a voice expressing thoughts than when they read the same thoughts in text. Adding human visual cues to text (i.e., seeing a person perform a script in a subtitled video clip), did not increase the likelihood of inferring a human creator compared with only reading text, suggesting that defining features of personhood may be conveyed more clearly in speech (Experiments 1 and 2). Removing the naturalistic paralinguistic cues that convey humanlike capacity for thinking and feeling, such as varied pace and intonation, eliminates the humanizing effect of speech (Experiment 4). We discuss implications for dehumanizing others through text-based media, and for anthropomorphizing machines through speech-based media.

Keywords: communication, voice, mind perception, dehumanization, anthropomorphism

Alan Turing (1950) created a famous benchmark for determining whether a computer can "think:" when it can convince a majority of people that they are interacting with a person instead of a machine. Turing's link between thinking and personhood is no accident (Farah & Heberlein, 2007). Boethus, writing in the 6th century, defined personhood as "an individual substance of rational nature" (Singer, 1994). Centuries later, John Locke (1841/1997) echoed this definition of a person as "an intelligent being that has reason and reflection." Immanuel Kant (1785/1993) used this definition of personhood as the guiding light of morality, noting that, "rational beings are called persons inasmuch as their nature already marks them out as ends in themselves." Recent surveys of laypeople likewise identify the capacity for thinking as a unique feature of humanity (Leyens et al., 2000, 2001). A person has a mind capable of thinking but a computer does not.

As clear as this reality may be, it may not be so clear psychologically. People sometimes recognize a thoughtful mind in their cars, computers, or other mindless gadgets (Guthrie, 1995; Naas, 2010). A robot that moves at a humanlike pace seems more thoughtful than a relatively sluggish or frantic robot (Morewedge, Preston, & Wegner, 2007). An autonomous automobile that interacts with you using a human voice while driving itself seems "smarter," and therefore, more trustworthy, than a noninteractive vehicle (Waytz, Heafner, & Epley, 2014). Attributing humanlike mental capacities of thinking and feeling to nonhuman agents is the essence of anthropomorphism (Epley, Waytz, & Cacioppo, 2007).

Inversely, people sometimes fail to recognize a thoughtful mind in other human beings, treating them instead like relatively mindless animals or objects (Haslam, 2006). Failing to attribute a humanlike mind to another person is the essence of dehumanization. These twin phenomena of anthropomorphism and dehumanization raise a fundamental question in social life that goes far beyond Turing's test: how does an agent convey the fundamentally humanlike capacity to think or feel? Answering this question will predict when machines might be treated like people, and when people might be treated like machines. It also predicts when machines might be mistaken for people, and people mistaken for machines.

Existing theory predicts that anthropomorphism and dehumanization are determined by features of the agent being perceived (e.g., morphology, motion, and observed behavior) and by features

Juliana Schroeder, Haas School of Business, University of California, Berkeley; Nicholas Epley, Booth School of Business, University of Chicago.

We thank the Neubauer Family Faculty Fellowship for financial support of this research, and Daniel Gilbert for suggesting the main dependent variable used in these experiments. We also thank Jasmine Kwong, Michal Dzitko, Annette Felton, Shreya Kalva, Alex Kristal, Paul Lou, Adam Picker, Megan Porter, Sunni Rogers, Jenna Rozelle, Max Snyder, and Sherry Tseng for assistance conducting these experiments. Portions of this research were presented at the 2013 Annual Meeting of the Society for Personality and Social Psychology.

Correspondence concerning this article should be addressed to Juliana Schroeder, Haas School of Business, University of California, Berkeley, 2220 Piedmont Avenue, Berkeley, CA 94720. E-mail: jschroeder@haas.berkeley.edu

of the perceiving agent (e.g., group affiliation, motivation, and social connection; Epley, Waytz, & Cacioppo, 2007). Here we test what we believe is a particularly important feature of the agent being perceived: a humanlike voice. Beyond the semantic content in speech, a humanlike voice also conveys paralinguistic information (e.g., volume, tone, and rate) that provides additional insight into one's thoughts and feelings. Indeed, voice evolved in large part as a tool to communicate an agent's mind to others through speech (Pinker & Bloom, 1990), and people can more accurately estimate others' mental states when they hear someone speak than when they read the same words in text (Hall & Schmid Mast, 2007; Kruger, Epley, Parker, & Ng, 2005). Therefore, we predicted that communicating with a humanlike voice would make an agent seem more like a person (vs. machine) than communicating the same content through other communication media (e.g., reading text, observing body language, or speaking with a voice that lacks critical paralinguistic cues). Adding a humanlike voice to a machine might make it seem more like a person (i.e., anthropomorphism). Likewise, *removing* voice from an actual person by communicating through text might make a person seem more like a machine (i.e., dehumanization).

Several existing experiments suggest that adding a humanlike voice to computerized agents increases anthropomorphism (Nass & Brave, 2005; Takayama & Nass, 2008; Waytz, Heafner, & Epley, 2014). Our experiments go beyond these results by providing a more precise understanding of the interpersonal consequences of hearing a humanlike voice compared with observing other inferential cues (e.g., visual cues such as seeing a human), by identifying which cues in voice convey personhood, and by testing the inverse possibility that *removing* voice from human interactions might lead to dehumanized perceptions of a speaker. This latter hypothesis is especially important as technology makes text-based interactions increasingly common in everyday life.

We test how a humanlike speaking voice affects dehumanization and anthropomorphism in four experiments. For each experiment, we report how we determined our sample size, all data exclusions, all manipulations, and all measures. Data for all experiments can be retrieved at http://faculty.haas.berkeley.edu/jschroeder/data.html. Experiments 1 and 3 remove voice and measure dehumanization (mistaking a human for a computer) whereas Experiment 2 adds voice and measures anthropomorphism (mistaking a computer for a human). All experiments test the effect of hearing speech versus reading the same words in text. To increase generalizability and remove confounds, we generated text using different methods in each experiment: transcriptions of human speech (Exp. 1), computer-generated text (Experiment 2), or written essays (Experiments 3 and 4). We added a third channel of human visual cues in Experiments 1–3 to determine whether speech is uniquely humanizing, and to test an alternative explanation that any humanlike cue will reduce dehumanization and increase anthropomorphism. We predicted a speaker's voice would be uniquely humanizing because it contains paralinguistic cues that reveal active mental experiences of thinking and feeling, and that these cues uniquely reveal the presence of a humanlike mind (Schroeder & Epley, 2015).

Our final experiment tests why speech is humanizing. We suggest that paralinguistic cues in a person's voice can convey the presence of a lively, thoughtful, and active mind, in a way that is analogous to how visual cues convey the presence of biological life. A person can tell whether another agent is alive or dead because of variance in motion. A living person's body moves in naturalistic ways. A dead person's body is still, with no variance in motion. Likewise, a person's voice also contains naturalistic variance through paralinguistic cues that may analogously convey the presence of a lively and active mind. A speaker's pitch rises and falls over time, yielding variance in tone (intonation). A speaker's pace quickens and slows, producing variance in speech rate. The presence of a lively and active mind—a humanlike mind—could be reflected through these cues. A rising pitch, for instance, may reflect confidence in judgment whereas a falling pitch may reflect more careful deliberation. Consistent with this possibility, speakers communicate mental states such as the valence of emotional experience or intentions even when speech lacks any meaningful semantic content (McAleer, Todorov, & Belin, 2014; Scherer, Banse, & Wallbott, 2001; Weisbuch, Pauker, & Ambady, 2009). This predicts that a person with a speaking voice lacking naturalistic variance in paralinguistic cues would be evaluated similar to a person being evaluated over text or another communication media devoid of paralinguistic cues (such as silent video). A voice lacking variance in paralinguistic cues might make another person's mind seem relatively dead or dull, more like a mindless machine than like a mindful human being.

Experiment 1: Dehumanization

We examined our hypotheses by using the inverse of Alan Turing's (1950) famous test: are observers more likely to mistakenly believe genuine human speech was created by a computer when they hear the speech than when they read the very same words? To rule out the artifact that any human cue might lead observers to think the script was created by a human, we also added visual cues: seeing a human. If a person's voice uniquely conveys humanlike mental capacities, as we predict, then we should find an effect of voice, and no unique effect of visual cues, on judgments of a speech's creator. To test our prediction, we created a novel paradigm in which individuals evaluate whether the *creator* of the content that they hear or read was a computer or a human. Note that participants are not evaluating whether a voice is that of a computer or a human, but rather whether the *creator* of the content was a human or a machine. This is essential because computerized voices could obviously sound different than a real human voice. Our hypotheses are not testing how people evaluate voices in communication, but rather about the inferences people make about the agents who are actually communicating."

Method

Participants. We decided to collect at least 20 speakers to obtain a more ecologically valid range of stimuli than is common in most psychology experiments (Fiedler, 2011; Kenny, 1985; Wells & Windschitl, 1999). This stimulus sampling is an important feature of all of our experiments because it enables us to assess how people evaluate a range of naturalistic speakers that might be encountered in everyday life (Brunswik, 1947). It also eliminates concerns that any result might be produced by some idiosyncratic feature of a single person's voice, or by some artifact introduced by creating artificial speech content. Naturalistic stimulus sampling enables stronger inferences about the strength of a phenom-

enon in the midst of the perceptually rich and chaotic environment of everyday life.

We initially recruited 33 people (M_{age} = 20.0, SD_{age} = 2.3, 58% female) from a Chicago research laboratory to serve as speakers in exchange for $2.00. Each speaker created two videos: one talking about a positive emotional experience and the other talking about a negative emotional experience (in counterbalanced order). We had no a priori prediction about the effect of emotional valence; we manipulated valence only as a robustness check for the magnitude of our predicted effect of speech (vs. text). To obtain our target of 20 speakers, two independent raters subsequently coded the videos based on the extent to which the speaker followed our instructions. Specifically, they evaluated "the extent to which participants talked about a very emotional experience and described all of their emotions in detail" on a scale of 0 (*not at all*), 1 (*somewhat*), or 2 (*very much*), r = .73. We included the 20 speakers who followed our instructions the best on this measure across their two speeches, giving us a final sample of 40 total speeches (20 positive experiences and 20 negative experiences). The 20 speakers included in the study did not differ significantly from the 13 speakers not included based on their gender, $\chi^2(1, 33) = 0.12$, $p > .10$, or their age, $t(31) = 1.23$, $p > .10$.

We decided to collect at least 640 observers so that at least four would watch each type of stimulus for each of the 40 videos in the four experimental conditions (160 conditions total). In total, 652 observers (M_{age} = 32.0, SD_{age} = 10.6, 42% female, 7 missing gender) from Amazon.com's Mechanical Turk participated. These observers completed the experiment in exchange for $0.30. We removed five observers who did not complete our primary dependent variable from analysis.

Speaker procedure. Participants described both a positive and negative emotional experience in their life (in counterbalanced order) on videotape. We selected this speech topic because it is generally humanizing, to the extent that most people believe humans have more emotional experience than do machines. For the positive emotional experience, the experimenter asked participants to:

> Think back on an important positive emotional experience that you had. Talk about the entire experience and all of the positive emotions that you felt during the experience. Describe your emotions from beginning to end. This should be an experience that led you to feel very emotional, such as feeling deep happiness. Be sure while telling your story to explain all of your feelings and emotions throughout the entire experience.

Instructions for the negative emotional experience were modified to refer to experiences of "deep sadness" instead of "deep happiness." The experimenter asked the speakers to repeat the instructions to ensure that they understood. Speakers then sat in a chair facing a video camera. To make the speeches as natural as possible, the experimenter stood behind the camera and told speakers to look at the camera and pretend like they were talking directly to the experimenter. The speaker talked until he or she was finished telling the story. Speech lengths varied from 1 to 3 min. The experimenter then stopped the video camera and read the next set of instructions for the negative emotional experience video (order counterbalanced). Again, speakers repeated the instructions, sat facing the video camera, and gave their speech. Finally, the experimenter debriefed speakers.

One research assistant transcribed the speeches after the verbatim transcript procedure used in United States courtroom depositions. A second assistant checked the transcriptions for accuracy.

Observer procedure. We randomly assigned observers to one of four experimental conditions: listening to a speaker (audio condition), watching and listening to a speaker (audiovisual condition), reading the speech (text condition), or reading and watching a speaker (subtitled video condition, with no sound included). To make the presentation of stimuli as consistent as possible across conditions, we presented all stimuli to observers as videos. Observers in the audio and text conditions, therefore, saw a black video screen and either heard the words or read them on the screen, respectively. This paradigm allows us to keep constant the amount of time each observer spent on each stimulus.

Observers in the text condition read the following information before observing the stimuli: "As you may know, computer technology is now attempting to mimic real human speech. Some computer programs are good enough that they can convince some observers that they are real people, whereas others are not as good. For the next few minutes, you will read the text of a script. Your job is to figure out whether the content of this script was originally created by a computer trying to create a script that sounded like a real human being or whether it was created by an actual human." Observers in the audio version read the same instructions, except the third sentence read, "For the next few minutes, you will listen to an actor (or actress) reciting a script." Observers in the audiovisual and subtitled video condition likewise read the same instructions, except the third sentence read, "For the next few minutes, you will watch an actor (or actress) reciting a memorized script."

To make sure that observers understood their task, those in the audio condition received further clarification: "To be absolutely clear, your job is not to determine whether the voice is of a real person or not. You will hear a real human actor reading a script. Your job is to determine whether the script itself was originally written by a computer or an actual human." Observers in the audiovisual and subtitled video condition, in contrast, received this clarification: "To be absolutely clear, your job is not to determine whether the actor or actress is a real person or not. You will watch a real human actor reciting a memorized script. Your job is to determine whether the script itself was originally written by a computer or an actual human."

After reporting whether the script was originally created by a human or computer, participants also reported how confident they were that their answer was correct on an 11-point response scale (0 = *not at all confident*, 5 = *moderately confident*, 10 = *absolutely confident*), and then explained "why they made the choice they did" in a free-response box. We did not analyze the free responses. Participants reported their confidence in this same way in each of the following experiments as well (Experiments 2–4), but confidence did not vary reliably by condition in any experiment reported in the manuscript and we, therefore, do not discuss it further.

Results. Whether the speech was about a negative or positive emotional experience did not affect the human versus computer judgment, $F(1, 639) = 0.01$, nor did it interact with communication medium on this judgment, $F(3, 639) = 0.82$. Therefore, we collapsed across this factor in the following analyses.

Observers' judgments of the script's creator varied by experimental condition, $F(3, 643) = 10.76$, $p < .01$, $\eta^2 = .05$. Because we used a nested experimental design (multiple observers for each speaker), we analyzed the effect of each condition (fixed factors) in a hierarchical regression controlling for the effect of speaker (random factor). As shown in Figure 1, removing voice was dehumanizing: observers who read the speeches were less likely to believe it was created by a human (text condition; $M = 53.6\%$, $SD = 50.0\%$) than observers who listened to them ($M = 80.8\%$, $SD = 39.5\%$), $t(627) = 5.29$, $p < .01$, $d = 0.42$. Adding individuating visual cues to the voice in the audiovisual condition did not increase the percentage who guessed the script was created by a human (audiovisual condition; $M = 71.6\%$, $SD = 45.2\%$) compared with audio alone, $t(628) = -1.79$, $p = .07$, $d = 0.14$. In contrast, stripping away the person's voice while retaining text in the subtitled video condition reduced the percentage who guessed the script was created by a human (subtitled video condition; $M = 60.7\%$, $SD = 49.0\%$) compared with the audio condition alone, $t(628) = 3.70$, $p < .01$, $d = 0.29$. The text and subtitled video conditions did not differ significantly from each other, $t(628) = 1.43$, $p = .15$, $d = 0.11$. An observer was most likely to believe a script was created by a human when they heard the speaker's voice.

Discussion

Participants who heard a speech were more likely to believe its content was created by a human (vs. computer) than participants who read the very same speech. The cues in voice seem uniquely humanizing, as visual cues—being able to see speakers in addition to hearing their voices—did not increase the percentage of evaluators who believed the script was created by a human. Being able to see a speaker without hearing his or her voice, however, significantly decreased the percentage who believed the script was created by a human. The presence of a mindful human creator was most clearly conveyed through a person's voice rather than by the overall amount of individuating (humanlike) cues available.

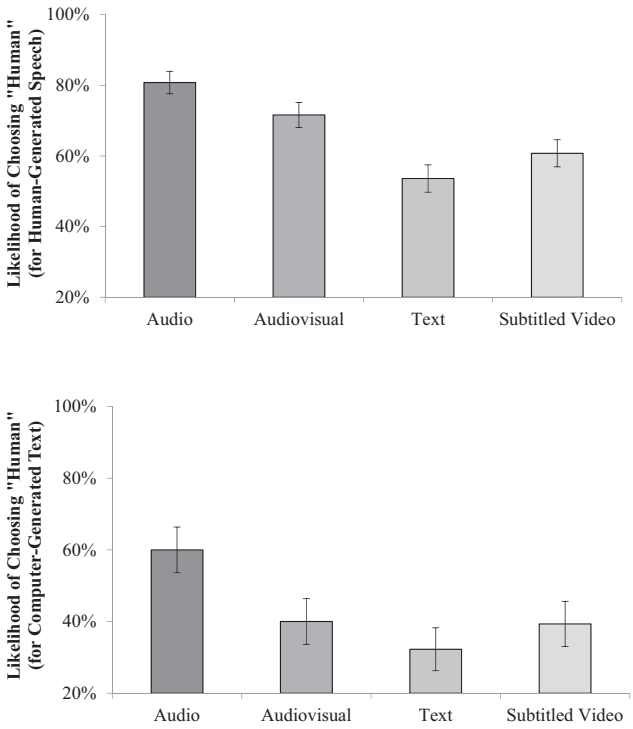

Figure 1. Percentage of observers in Experiment 1 (Panel 1; $n = 652$; stimuli created from human-generated speech) and Experiment 2 (Panel 2; $n = 243$; stimuli created from computer-generated text) who believed a script had been created by a human (vs. computer) in the audio, audiovisual, text, and subtitled video conditions. Errors bars represent the *SEM*.

Experiment 2: Anthropomorphism

The results of Experiment 1 suggest that removing a voice from human-generated speech can lead people to believe its content was created by a mindless machine. Experiment 2 tests the inverse: can adding a human voice to computer-generated speech lead people to believe its content was created by a mindful human being? This experiment is therefore closer to an actual Turing test, examining when human observers might mistake a mindless machine for a mindful person.

Method

Participants. We predetermined a sample size of at least 10 speakers and 240 observers. These sample sizes were smaller than Experiment 1 because speakers all read the same computer-generated text and so we expected less variability in evaluations across speakers. Ten people (50% female) from a Chicago research laboratory served as speakers in exchange for $3.00. Subsequently, 243 people ($M_{age} = 30.9$, $SD_{age} = 10.0$, 38% female) from Amazon Mechanical Turk served as observers in exchange for $0.30.

Speaker procedure. We created the essay using a "Postmodernism Generator" (http://www.elsewhere.org/pomo/) that uses a computer system to generate random text from recursive grammars (Bulhak, 1996). The full text was:

> "Society is elitist," says Derrida. It could be said that Porter suggests that we have to choose between material nihilism and neocultural theory. "Truth is intrinsically dead," says Debord; however, according to Reicher, it is not so much truth that is intrinsically dead, but rather the paradigm, and subsequent stasis, of truth. Sartre promotes the use of dialectic subconceptualist theory to deconstruct the status quo. However, Baudrillard uses the term 'cultural feminism' to denote the common ground between sexual identity and class. Marx suggests the use of material nihilism to analyze sexual identity. But Bataille uses the term 'cultural feminism' to denote a self-supporting totality. The example of material nihilism intrinsic to Gibson's Count Zero emerges again in Virtual Light, although in a more mythopoetical sense. Therefore, Baudrillard promotes the use of subdeconstructivist theory to challenge hierarchy. Debord's model of semantic Marxism states that culture is part of the futility of consciousness. It could be said that the main theme of the works of Gibson is the failure, and hence the dialectic, of postcultural society. Cultural feminism holds that context comes from the masses. But Bataille suggests the use of Lyotardist narrative to read and analyze art.

Just before giving their speech, speakers read the following instructions that were intended to maintain the natural paralinguistic cues present in actual human speech:

> When you perform, please imagine that you are the person who wrote the essay. Imbue your words with all of the thoughts, emotions, and

> substance that the writer him/herself felt. Read it as if you were actually coming up with the lines naturally off the top of your head rather than reading from an essay. Speak as naturally as you would if you were in the midst of a real conversation.

Observer procedure. We created the same four types of stimuli described in Experiment 1: (a) audio, (b) audiovisual, (c) text, and (d) subtitled video. Observers watched one of these four stimuli, then guessed whether the script was "originally created by a human or computer."

Results

Because observers were not fully nested within speakers (i.e., observers who read the text always read the same essay), we did not analyze the effects using hierarchical models. Observers' judgments of the script's creator varied as predicted by experimental condition, $F(3, 239) = 3.62$, $p = .01$, $\eta^2 = .04$. As shown in Figure 1, adding voice increased the tendency to anthropomorphize computer-generated text. Observers who listened to the speeches (audio condition) were more likely to guess the script was created by a human ($M = 60.0\%$, $SD = 49.4\%$) than did those who read identical speeches (text condition; $M = 32.3\%$, $SD = 47.1\%$), $t(239) = 3.14$, $p < .01$, $d = 0.41$. Adding visual human cues to the voice in the audiovisual condition did not significantly increase the percentage who believed the text was created by a human ($M = 40.0\%$, $SD = 49.4\%$) compared to the audio condition. In fact, it unexpectedly decreased it, $t(239) = -2.25$, $p = .03$, $d = 0.29$. Stripping away a person's voice while retaining text in the subtitled video condition also significantly reduced the percentage who believed the script was created by a human ($M = 39.3\%$, $SD = 49.3\%$) compared with the audio condition, $t(239) = 2.33$, $p = .02$, $d = 0.30$. Human versus computer judgments did not vary between the audiovisual, subtitled video, and text conditions, ts < 1.

Discussion

Participants in two experiments were more likely to believe that the content of a speech was created by a mindful human being rather than by a mindless computer when they heard a speech than when they read the identical content in text. This result did not simply reflect an increase in accuracy for identifying authentic human content through a person's voice, but rather a systematic bias to infer that a speech was created by a human after hearing it spoken.

Adding individuating cues to a person's voice through video again did not increase the tendency to guess that a script was created by a human. If anything, watching someone recite computer-generated text reduced the tendency to guess the text was created by a person compared with audio alone. We did not predict this reduction, do not observe it elsewhere, and do not speculate on it further. We note simply that these results again suggest that a humanlike voice may be uniquely equipped for conveying the presence of a humanlike mind.

Experiment 3: Job Candidates

Experiments 1 and 2 suggest that text is subtly dehumanizing, making a creator seem less likely to be a human regardless of whether the creator was actually a human or a computer. Creating these stimuli required either removing a human voice from speech (creating a transcript), or adding it to computerized text. One condition was therefore always derived from another, creating a potentially problematic confound in our design. In Experiment 3, human participants generated *both* spoken and written stimuli, removing this confound. Because semantic content may systematically differ in spoken and written communication, we also asked observers to evaluate transcriptions of participants' speeches. This enables two tests of our voice hypothesis, one comparing spoken versus written content that may vary in semantic content, and another comparing spoken content versus a transcript that contains identical semantic content.

Experiment 3 also tests our hypotheses in a context of practical relevance: short "elevator pitches" to a potential employer explaining a candidate's qualifications for the job. In prior research, we found that hypothetical job candidates were judged to be less thoughtful, less intelligent, and less worthy of being hired when potential employers read a transcript or written elevator pitch compared with hearing the pitch (Schroeder & Epley, 2015). Because the capacity for thinking is a defining feature of personhood, we expected that observers would also believe that the content of an elevator pitch was less likely to have been created by a real person when they read what a candidate has to say than when they hear it.

Method

Participants (speakers and observers). Speakers were 18 University of Chicago Booth School of Business students ($M_{age} = 28.2$, $SD_{age} = 2.07$, 39% female) who responded to our request for research assistance. Speakers received a $5 Starbucks gift card as compensation. These stimuli are also used in Schroeder and Epley (2015).

We decided to collect at least 270 observers so that at least five would watch each type of stimulus for each of the 18 videos in the three experimental conditions (54 conditions total). In total, 273 people ($M_{age} = 34.8$, $SD_{age} = 13.5$, 50% female) from the Museum of Science and Industry in Chicago served as speakers in exchange for a small snack.

Speaker procedure. We recruited MBA students to participate as job candidates in a study on how people make hiring decisions. Candidates first named the company for which they would most like to work, and then considered the pitch they would make to encourage this company to evaluate them positively and to hire them. Candidates made both a spoken and a written pitch to prospective employers (order counterbalanced). In the spoken pitch condition, we told candidates we would videotape them as they gave their pitch and that they should speak directly to the camera. We told candidates they had 2 min to talk but allowed them to reach the natural conclusion of their pitch (actual videos times ranged from 49 s to 2 min and 30 s). In the written pitch condition, we told candidates to compose a letter to a prospective employer. Candidates had 10 min to type their letter on a computer, after which we told them to finish their thought and stop typing. We arrived at these suggested time limits by asking two MBA students to create spoken and written pitches without any time restrictions, and then timed how long it took for them to speak or write their pitches. After finishing both their spoken and written

pitches, candidates completed a short survey. These items are unrelated to the current hypotheses.

To create transcripts of the spoken job pitches, one research assistant transcribed the spoken pitches and a second checked them for accuracy. To make the files more readable, we removed verbal filler words unless their exclusion changed the sentence's meaning.

Observer procedure. Observers were visitors from the Museum of Science and Industry in Chicago who agreed to serve as hypothetical employers. Observers consented to take a short survey in which they would decide whether a job pitch was created by a human or computer. We randomly assigned observers to one of three experimental conditions: listening to a spoken pitch (audio condition), reading a transcript of a spoken pitch (transcript condition), or reading the written pitch (writing condition). We gave observers the same instructions described in Experiment 1. After observing the stimuli, participants then reported whether the job pitch was "originally created by a human or computer."

Results

Observers' judgments of the job pitch's creator varied by experimental condition, $F(2, 270) = 6.18$, $p < .01$, $\eta^2 = .04$. Because we used a nested experimental design (multiple observers for each speaker), we analyzed the effect of each condition (fixed factors) in a hierarchical regression controlling for the effect of speaker (random factor). As shown in Figure 2, voice was humanizing: observers who listened to the pitches (audio condition) were more likely to believe it was created by a human ($M = 79.1\%$, $SD = 40.9\%$) than did those who read the identical text from these speeches (transcript condition; $M = 55.9\%$, $SD = 49.9\%$), $t(270) = 3.34$, $p < .01$, $d = 0.41$, or those who read the written pitches (written condition: $M = 59.6\%$, $SD = 49.3\%$), $t(270) = 2.74$, $p = .01$, $d = 0.33$). There was no difference in beliefs about the pitch's creator in the transcript versus written conditions, $t < 1$.

Discussion

Hypothetical employers who listened to a job candidate's elevator pitch explaining his or her qualifications believed the content was more likely to have been created by a machine when they read what the candidate had to say—either in a written pitch or a transcript—than when they heard what the candidate had to say. These results converge with prior research (Schroeder & Epley, 2015) to suggest that a person's voice conveys the fundamental human capacity for thinking and reasoning. Lacking those cues from a person's voice in text, readers do not seem to add them in spontaneously while reading someone's text, in this case decreasing the perceived likelihood that a written elevator pitch was created by an actual person.

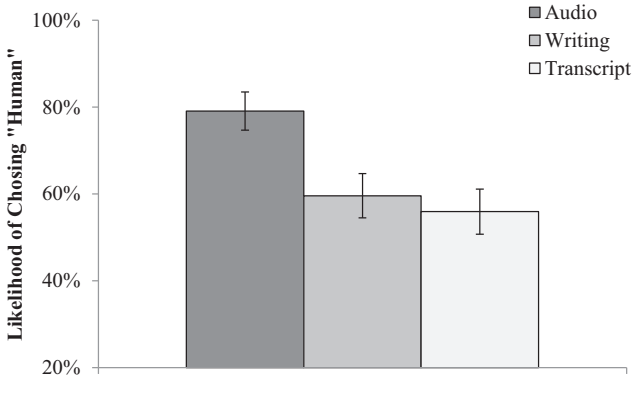

Figure 2. Percentage of observers in Experiment 3 ($n = 273$) who believed a script had been created by a human (vs. computer) in the audio, writing, and transcript conditions. Errors bars represent the *SEM*.

Although Experiments 1–3 suggest that voice moderates anthropomorphism and dehumanization, they do not explain which features of a person's voice do so. Earlier we hypothesized that a person's voice may contain cues to mental life in much the same way an agent's body contains cues to biological life. In particular, biological life is revealed through bodily motion. Lacking any motion or movement, an agent appears to be either dead or sleeping. Likewise, a person's voice also contains motion through variance in pitch (intonation) and pace, among others, providing cues to a mental life of conscious thought or emotional experience (Gray & Wegner, 2008; Morewedge et al., 2007; Schroeder & Epley, 2015). If these cues convey the mental capacity for thinking and feeling, and hence personhood, then voice lacking these cues should lead to the same inferences as observers reading text alone. We tested this directly in Experiment 4.

Experiment 4: Mindless Voice

We examined the mechanism by which a person's voice conveys the presence of a humanlike mind by manipulating the paralinguistic cues (e.g., volume, pitch, and rate of speech) that a human voice adds beyond the semantic content of language, and then testing whether the objective qualities of these cues mediated the influence of a speaker's voice on observers' inferences. Observers therefore listened to either a "humanlike" voice that contained the natural paralinguistic cues that a person would typically use when expressing himself or herself, or they listened to a voice that lacked these cues. Our theory is that paralinguistic cues are closely connected to actual conscious thought processes as they occur, thereby serving as an "honest" cue to the presence of mental life. Naturalistic intonation (typical human variance in pitch) may reflect emotional experience or rational thought. Changes in volume may likewise reflect emotional experience or certainty in judgment. If so, then removing these cues from a person's voice should make him or her seem less humanlike, just as we observed with text-based communication.

Method

Participants. In line with the preceding experiments, we intended to collect between 10 and 20 communicators. Fifteen people ($M_{age} = 21.8$, $SD_{age} = 3.4$, 53% female) from a Chicago research laboratory served as writers in exchange for $4.00. We then recruited four actors from a University drama department (2 male, 2 female, $M_{age} = 20$) to serve as speakers in exchange for $25.00. We predetermined a sample of 300 observers (at least 12 observers in each of 25 experimental conditions) to ensure sufficient statistical power given the additional variability we expected because of using different writers and different speakers. Adults visiting the Museum of Science and Industry in Chicago served as observers ($n = 300$, $M_{age} = 32.2$, $SD_{age} = 12.1$, 48% female) in exchange for a small snack.

Writer procedure. Writers wrote a description (using a computer) of an important life decision that ended poorly and one that ended well in counterbalanced order. To select essays that could conceivably be created by either a human or a computer (thereby avoiding ceiling or floor effects on our dependent variable), we asked 325 Amazon Mechanical Turk workers to indicate whether they believed the essay was created by a human or a computer. The ratings were normally distributed, ranging from 30% to 92% who believed the script was created by a human. We selected five essays closest to the median estimate of this distribution (60%) to use as stimuli.

Speaker procedure. Actors read the selected essays in a sound booth. Male actors read the three essays originally written by men, whereas female actors read the two essays originally written by women. To create stimuli that contained a more or less humanlike voice, each actor read a given essay twice, first in a "mindful" voice using the same instructions provided to speakers in Experiment 2, and then in a "mindless" voice that asked speakers to "read the words exactly as you see them on the page. Put little feeling or life into the words."

Observer procedure. Observers listened to an actor reading the statement in a mindful voice ($n = 120$) or a mindless voice ($n = 120$), or read the original essay ($n = 60$). We collected double the number of participants in the voice conditions because two actors read each essay, providing an equal number of participants evaluating each actor's reading of the essay as in the text condition. We told observers in the two audio conditions that an actor was reading from the original written essay. After reading or listening to the essay, observers judged whether the essay was originally created by a human or a computer.

Results

As shown in Figure 3, more observers guessed that the essay was created by a human when they listened to the mindful voices ($M = 65.0\%$, $SE = 4.6\%$) than when they read the text ($M = 47.2\%$, $SE = 6.5\%$), $\chi^2(1,180) = 5.55$, $p = .02$, $\varphi = 0.18$, or listened to mindless voices ($M = 50.3\%$, $SE = 4.6\%$), $\chi^2(1,240) = 5.52$, $p = .02$, $\varphi = 0.15$. The text and mindless voice conditions did not differ from each other, $\chi^2(1,180) = 0.18$, $p > .10$. These results demonstrate that altering the speaker's voice altered perceptions of the content's creator.

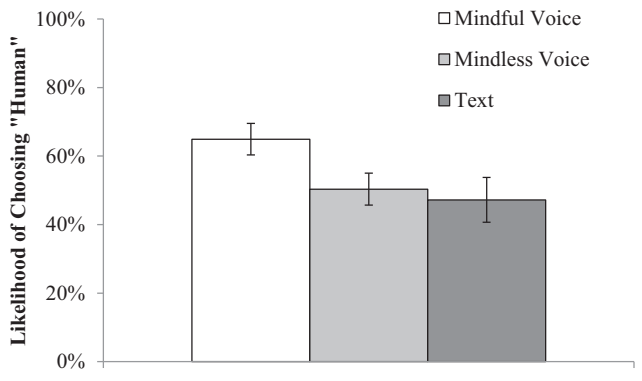

Figure 3. Percentage of observers in Experiment 4 ($n = 300$) who correctly guessed that a script had been created by a human (vs. computer) in the mindful voice, mindless voice, and text conditions. Errors bars represent the *SEM*.

Differences in paralinguistic cues must account for these results because the semantic content was identical across conditions. Based on prior literature examining how paralinguistic cues affect trait-based impression (e.g., Addington, 1968; Collins & Missing, 2003; Gregory & Webster, 1996; Hughes, Mogilski, & Harrison, 2014; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Laplante & Ambady, 2003; Niedzielski, 1999), we extracted the following cues using Praat software (Boersma & Weenink, 2016): mean pitch, mean amplitude, *SD* of pitch, *SD* of amplitude, speech length, and mean pause length for each of the 20 voice clips.[1] Our purpose was to test whether any of these cues mediated the effect of mindful versus mindless voice on perceptions of the content's creator.

As intended, the mindful voices significantly differed on almost all of these paralinguistic cues compared to their mindless voices. Mindful voices had higher mean pitch ($M = 165.5$ Hz, $SD = 50.5$ vs. $M = 147.7$ Hz, $SD = 48.4$), $t(9) = 7.76$, $p < .01$, $d = 2.45$, higher *SD* of pitch ($M = 38.8$ Hz, $SD = 6.6$ vs. $M = 23.7$ Hz, $SD = 8.7$), $t(9) = 9.51$, $p < .01$, $d = 3.01$, higher mean amplitude ($M = 71.4$ dB, $SD = 5.7$ vs. $M = 62.2$ dB, $SD = 7.0$), $t(9) = 3.39$, $p = .01$, $d = 1.07$, higher *SD* of amplitude ($M = 9.3$ dB, $SD = 1.4$ vs. $M = 7.6$ dB, $SD = 1.2$), $t(9) = 8.94$, $p < .01$, $d = 2.83$, and marginally longer overall length ($M = 60.9$ s, $SD = 11.1$ vs. $M = 55.7$ s, $SD = 10.1$), $t(9) = 1.97$, $p = .08$, $d = 0.62$. We observed no difference in the mean pause length, $t(9) < 1.37$, of mindful versus mindless voices.

To identify which of these paralinguistic cues predicted observers' estimates of the speech's creator, we conducted a linear regression using each of these paralinguistic cues as independent variables. Only intonation (the *SD* of pitch) significantly predicted evaluations in this regression, $\beta = .24$, $p < .01$. When we included intonation in a bootstrapped regression model predicting the effect of experimental condition on judgments, the effect of voice condition became nonsignificant (from $\beta = .62$, $SE = .26$, $p = .02$, to $\beta = .20$, $SE = .33$, $p = .54$), but the effect of intonation remained significant ($\beta = .03$, $SE = .01$, $p = .03$), demonstrating that intonation fully mediated the effect of the audio condition on observers' judgments. A 5,000-sample bootstrap test estimated a standardized indirect effect of .44 ($SE = .21$, 95% bias-corrected CI [.02, .86]), indicating a significant indirect effect (MacKinnon, Fairchild, & Fritz, 2007).

Discussion

Observers who listened to an actor speaking mindfully, as if actually experiencing the thoughts and feelings described in an essay were more likely to believe the essay was created by a human (vs. computer) compared with observers reading the essay

[1] To compute each speaker's pitch profile in Praat, we first set a fixed time step of 0.01 s. We set the pitch range to 150 to 500 Hz for female speakers and 75 to 500 Hz for male speakers. We used the autocorrelation analysis method in the pitch settings. To export the pitch, we selected the entire pitch profile and saved the pitch listing as a text file that we imported into Excel. We then computed the average and *SD* of the pitch in Excel. To compute amplitude, we used the default Praat settings for the intensity analysis. We selected the intensity profile and imported it into Excel using the same method as we did for pitch. The number of seconds provided in these pitch and amplitude profiles composed our measure of speech length. To compute pause length, we counted the number of blank cells in the pitch profile (each Excel cell represented 0.01 s).

directly or those listening to the same actors speaking mindlessly, without any "feeling or life" in their voices. Subsequent analyses of paralinguistic cues suggested that intonation was the most important vocal cue for revealing the presence of a human mind. Altering these paralinguistic cues by asking actors to speak "as if only reading words on a page," resulted in significantly more monotone speeches, and yielded evaluations that did not differ from evaluations of the text alone.

This indirectly suggests that text may lead to dehumanized perceptions in our experiments because observers do not spontaneously add cues to mental life, such as intonation, when reading someone else's speech. The voice in one's head while reading may sound more monotone, and hence more robotic, than the voice of an actual human speaker.

General Discussion

A person has a mind capable of thinking and feeling but a computer does not. Four experiments suggest that this defining feature of personhood is communicated at least partly through the paralinguistic cues contained in speech, such that adding a humanlike voice to computer-generated text increased the tendency to infer that it was actually created by a real person (Experiment 2). In contrast, removing a voice from human-generated speech increased the tendency to presume the content was actually created by a computer (Experiment 1). In prior research (Schroeder & Epley, 2015), job candidates were judged to be more intelligent, thoughtful, and hirable when potential employers heard their elevator pitch than when they read the same text or read a written pitch. In Experiment 3, observers inferred that these elevator pitches were more likely to have been created by a real human being when observers heard what the candidate had to say than when they read what the candidate had to say. These results suggest that speech can be humanizing and its absence, dehumanizing.

These results appear to stem from the paralinguistic cues present in speech, especially intonation (i.e., variability in pitch, Experiment 4). Just as variability in bodily movement (i.e., biological motion) serves as a cue for biological life, our experiments suggest that variability in pitch (i.e., intonation) can serve as a cue to mental life, and hence the presence of a humanlike mind. Text alone does not contain these paralinguistic cues. If readers do not spontaneously imbue a passage of text with such cues as they are reading it, then they may be less likely to recognize the humanlike mind behind the words they read. Consistent with this possibility, observers in Experiment 4 who listened to actors read text in a lifeless way—without the intonation present in naturalistic human speech—evaluated them the same as observers who read the writer's text alone.

Existing research suggests that speech is uniquely equipped to convey a person's mental states and capacities. Observers judge a target's thoughts and feelings more accurately when they hear someone speak than when they read the same text alone (Kruger et al., 2005), or when they watch a person (silently) speaking (Gesn & Ickes, 1999; Hall & Schmid Mast, 2007; Kruger et al., 2005; Zaki et al., 2009). Most relevant for our current findings, adding an authentic humanlike voice to a mindless machine can increase the tendency to anthropomorphize it (Nass & Brave, 2005; Takayama & Nass, 2008; Waytz, Heafner, & Epley, 2014).

Our research advances existing research in two important ways. First, we provide a more nuanced comparison among different forms of communication, including both verbal and visual cues. In two experiments that directly compared the effect of adding a speaker's voice to semantic content (i.e., speech) compared to adding visual content (i.e., video), we observed that adding voice significantly affected observers' evaluations but that adding visual content did not (Experiments 1 and 2). We believe these results are interesting, but also inconclusive. On the one hand, these results could indicate an interesting fact about social cognition: that another person's humanlike mental capacities are heard more easily than they are seen. Nobody would attempt to teach the basic ideas of statistics through their facial expressions, or wax nostalgic about a family vacation through pantomime. Vocal cues might simply provide a much stronger signal for sophisticated thoughts and feelings than visual cues. On the other hand, these results could also indicate an uninteresting artifact of our experimental designs: subtitled video could have distracted viewers from nonverbal behavior, or the speeches we asked participants to give may not have allowed for much physical nonverbal behavior. Determining the relative importance of visual versus verbal cues in mind perception would require using stimuli that vary visual cues as clearly as our current experiments vary vocal cues. Understanding how different cues reveal different aspects of an otherwise hidden mind is a promising avenue for future research.

Second, we provide a more precise understanding of why speech might be humanizing: intonation (i.e., pitch variance) reveals a humanlike mind. Whereas other research has examined how mean level pitch affects *trait*-based evaluations of others (Addington, 1968; Collins & Missing, 2003; Feinberg et al., 2008; Gregory & Webster, 1996; Hughes et al., 2014; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010; Laplante & Ambady, 2003; Niedzielski, 1999; Ray & Ray, 1990; Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012), our results suggest that *variability* in pitch may convey the existence of humanlike mental capacities, leading observers to infer a human source. People naturally modulate their pitch when expressing thoughts; pitch rises when asking a question, falls when expressing sadness, and generally fluctuates as people reason and express ideas with enthusiasm and emotion.

The influence of intonation also helps to explain why text could be relatively dehumanizing: readers may not spontaneously infuse written content with the cues that would reveal a humanlike mind. This does not mean that a skilled author would be unable to humanize their writing, but rather that randomly selected readers do not spontaneously humanize text as they are reading it. Identifying precisely why voice is humanizing is important because it demonstrates how different types of voices might be evaluated systematically differently, regardless of whether the voices are coming from a real human being or from a machine. For computer scientists and engineers interested in humanizing technology even further, Experiment 4 suggests that accurately mimicking naturalistic intonation should be an especially important goal.

These results also have interesting implications for how people interact with both machines and human beings in the modern world. We consider each in turn.

Anthropomorphizing With Voice

Many technological devices now come equipped with humanlike speech. GPS systems direct drivers using computer-generated speech. Call centers are staffed by computers that talk to angry customers. And cell phone "assistants" ask, "What can I do for you?" Our experiments suggest that humanlike speech may be an especially important attribute of anthropomorphized machines. Our results are consistent with other correlational data linking the presence of a human voice to greater anthropomorphism of machines (Nass & Brave, 2005; Takayama & Nass, 2008). They are also consistent with a recent experiment testing the causal influence of an agent's voice on anthropomorphism. In this experiment, participants were randomly assigned to drive either a normal vehicle, an autonomous vehicle, or an "anthropomorphized" autonomous vehicle that had a real human voice rated the anthropomorphized vehicle as the most mindful and trusted it most to drive safely (Waytz et al., 2014). Just as in our experiments, a car with a voice of its own seems better able to think and feel than any normal car, and is one you might be willing to trust with your own life.

Neither this experiment, nor those we report in this paper, specifically identify the importance of voice in these important outcomes of anthropomorphism. The Waytz et al. (2014) experiment necessarily confounded the presence of thoughtful content with the presence of a human voice (because the voice actually included intelligent content). We did not measure the outcomes of anthropomorphizing machines in the present manuscript. Identifying the relative importance of voice compared to other attributes, such as content, is therefore an important direction for future research.

Our experiments suggest that intonation—pitch variance—may be a critical cue for humanizing these gadgets through their voices. Current technology appears limited in this regard compared with real voices. To examine the intonation in computerized voices, we created Siri voices (using either the "Alex" and "Kathy" Siri voice templates on Macintosh computers, matching genders) from the transcripts of the speakers talking about their emotional life experiences in Experiment 1, and compared those voices to the actual participants describing their experiences. The Siri voices had significantly less intonation than real human voices (M_{SIRI} = 24.34 Hz, SD = 3.80, vs. M_{Human} = 36.06 Hz, SD = 19.64), $t(39) = 3.97$, $p < .01$, $d = 0.83$. As people's reliance on technology continues to deepen in modern life, engineers might do well to understand the psychological processes that guide social cognition. Seemingly subtle cues such as intonation could have an effect on how people think about, and treat, their technology.

Dehumanizing Without Voice

Adding voice might anthropomorphize a phone, but our experiments demonstrate that removing the voice from a real person might be subtly dehumanizing, making the person seem more like a mindless machine. Our results have potential importance for everyday human interactions. We end by considering two contexts in particular: courtrooms and text-based social media.

In courtrooms, witnesses provide testimony that is transcribed verbatim by a stenographer. When deliberating jurors ask for more information about a person's testimony, they do not hear a recording of the testimony but instead get the verbatim transcript. Our experiments suggest that this seemingly innocuous procedure could have profound effects on how jurors evaluate a given piece of testimony. Was the victim truly remorseful? Was the crime a well-reasoned plan or an unintentional accident? How much did the victim really suffer? Many legal judgments rest on inferences about the minds of perpetrators and victims, inferences that depend on the medium through which information is conveyed.

More widely in everyday life, social interactions are becoming increasingly text-based, with people keeping in touch through technology such as texting or emailing rather than through talking. In 2012, a representative sample of 2,000 American adults reported spending 39% *more* time socializing online than face-to-face (Marketwired, 2012). Some researchers have suggested that these changing interaction norms are partly responsible for cross-generational changes, such as increases in narcissism, loneliness, and social apathy (Cacioppo & Patrick, 2008; Konrath, 2012; Turkle, 2012; Twenge, 2013). Our experiments suggest that these voiceless media might remove an important human element from these interactions. In one experiment, adolescent women facing a stressful event became more relaxed when they were able to call their mother on the phone and hear her voice directly, but became no more relaxed when they were able to "chat" using text with their mothers (Seltzer, Prososki, Ziegler, & Pollak, 2012). Although people are generally much happier connecting with others than being alone (Kahneman & Deaton, 2010), connecting with others online (using Facebook) in one study significantly reduced happiness over time (Kross et al., 2013). In another experiment, those randomly assigned to reconnect with an old friend over the telephone reported feeling more connected than those who reconnected over email (Kumar & Epley, 2016).

Technology enables connections between people at great distances, over the Internet, through text messages, via email, and over the phone. Emerging text-based media provide great technological gains in efficiency over speech-based media. Our research suggests these technological gains in efficiency, however, may come with some surprising psychological costs.

References

Addington, D. W. (1968). The relationship of selected vocal characteristics to personality perception. *Speech Monographs, 35,* 492–503. http://dx.doi.org/10.1080/03637756809375599

Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer [Computer program]. Retrieved from http://www.praat.org/

Brunswik, E. (1947). *Systematic and representative design of psychological experiments*. Berkeley, CA: University of California Press.

Bulhak, A. C. (1996, April). On the simulation of postmodernism and mental debility using recursive transition networks. Technical Report No. 96/264, Department of Computer Science, Monash University, Melbourne, Australia.

Cacioppo, J. T., & Patrick, B. (2008). *Loneliness: Human nature and the need for social connection*. New York, NY: Norton.

Collins, S. A., & Missing, C. (2003). Vocal and visual attractiveness are related in women. *Animal Behaviour, 65,* 997–1004. http://dx.doi.org/10.1006/anbe.2003.2123

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

Farah, M. J., & Heberlein, A. S. (2007). Personhood and neuroscience: Naturalizing or nihilating? *The American Journal of Bioethics, 7,* 37–48. http://dx.doi.org/10.1080/15265160601064199

Feinberg, D. R., DeBruine, L. M., Jones, B. C., & Little, A. C. (2008). Correlated preferences for men's facial and vocal masculinity. *Evolution and Human Behavior, 29,* 233–241. http://dx.doi.org/10.1016/j.evolhumbehav.2007.12.008

Fiedler, K. (2011). Voodoo correlations are everywhere: Not only in neuroscience. *Perspectives on Psychological Science, 6,* 163–171. http://dx.doi.org/10.1177/1745691611400237

Gesn, P. R., & Ickes, W. (1999). The development of meaning contexts for empathic accuracy: Channel and sequence effects. *Journal of Personality and Social Psychology, 77,* 746–761. http://dx.doi.org/10.1037/0022-3514.77.4.746

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science, 19,* 1260–1262. http://dx.doi.org/10.1111/j.1467-9280.2008.02208.x

Gregory, S. W., Jr., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology, 70,* 1231–1240. http://dx.doi.org/10.1037/0022-3514.70.6.1231

Guthrie, S. E. (1995). *Faces in the clouds: A new theory of religion*. New York, NY: Oxford University Press.

Hall, J. A., & Schmid Mast, M. (2007). Sources of accuracy in the empathic accuracy paradigm. *Emotion, 7,* 438–446. http://dx.doi.org/10.1037/1528-3542.7.2.438

Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review, 10,* 252–264. http://dx.doi.org/10.1207/s15327957pspr1003_4

Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2014). The perception and parameters of intentional voice manipulation. *Journal of Nonverbal Behavior, 38,* 107–127. http://dx.doi.org/10.1007/s10919-013-0163-z

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C., & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour, 79,* 57–62. http://dx.doi.org/10.1016/j.anbehav.2009.10.003

Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences of the United States of America, 107,* 16489–16493. http://dx.doi.org/10.1073/pnas.1011492107

Kant, I. (1993). Grounding for the Metaphysics of Morals, 3rd Ed. (J. W. Ellington, Trans.) Indianapolis, IN: Hackett Publishing Company.

Kenny, D. A. (1985). Quantitative methods for social psychology. In G. Lindzey & E. Aronson (Eds.), *Handbook of social psychology* (3rd ed., Vol. 1, pp. 487–508). New York, NY: Random House.

Konrath, S. (2012). The empathy paradox: Increasing disconnection in the age of increasing connection. In R. Luppicini (Ed.), *Handbook of research on technoself: Identity in a technological society* (pp. 204–228). Hershey, PA: IGI Global.

Kross, E., Verduyn, P., Demiralp, E., Park, J., Lee, D. S., Lin, N., . . . Ybarra, O. (2013). Facebook use predicts declines in subjective well-being in young adults. *PLoS ONE, 8,* e69841. http://dx.doi.org/10.1371/journal.pone.0069841

Kruger, J., Epley, N., Parker, J., & Ng, Z. W. (2005). Egocentrism over e-mail: Can we communicate as well as we think? *Journal of Personality and Social Psychology, 89,* 925–936. http://dx.doi.org/10.1037/0022-3514.89.6.925

Kumar, A., & Epley, N. (2016). *It's surprisingly nice to hear from you: Talking increases social connection compared to typing*. Unpublished data, University of Chicago.

Laplante, D., & Ambady, N. (2003). On how things are said: Voice tone, voice intensity, verbal content, and perceptions of politeness. *Journal of Language and Social Psychology, 22,* 434–441. http://dx.doi.org/10.1177/0261927X03258084

Leyens, J. P., Paladino, P. M., Rodriguez-Torres, R., Vaes, J., Demoulin, S., Rodriguez, A. P., & Gaunt, R. (2000). The emotional side of prejudice: The attribution of secondary emotions to ingroups and outgroups. *Personality and Social Psychology Review, 4,* 186–197. http://dx.doi.org/10.1207/S15327957PSPR0402_06

Leyens, J.-P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M.-P., Vaes, J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology, 31,* 395–411. http://dx.doi.org/10.1002/ejsp.50

Locke, J. (1997). *An essay concerning human understanding*. Harmondsworth, England: Penguin Books. (Original work published 1841)

MacKinnon, D. P., Fairchild, A. J., & Fritz, M. S. (2007). Mediation analysis. *Annual Review of Psychology, 58,* 593–614. http://dx.doi.org/10.1146/annurev.psych.58.110405.085542

Marketwired. (2012). *Generation lonely? 39 percent of Americans spend more time socializing online than face-to-face*. Retrieved from http://www.marketwired.com/press-release/-1648444.htm

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? Personality impressions from brief novel voices. *PLoS ONE, 9,* e90779. http://dx.doi.org/10.1371/journal.pone.0090779

Morewedge, C. K., Preston, J., & Wegner, D. M. (2007). Timescale bias in the attribution of mind. *Journal of Personality and Social Psychology, 93,* 1–11. http://dx.doi.org/10.1037/0022-3514.93.1.1

Nass, C. (2010). *The Man Who Lied to His Laptop: What Machines Teach Us About Human Relationships*. New York, NY: Penguin.

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

Niedzielski, N. (1999). The effect of social information on the perception of sociolinguistic variables. *Journal of Language and Social Psychology, 18,* 62–85. http://dx.doi.org/10.1177/0261927X99018001005

Pinker, S., & Bloom, P. (1990). Natural language and natural selection. *Behavioral and Brain Sciences, 13,* 707–727. http://dx.doi.org/10.1017/S0140525X00081061

Ray, E. B., & Ray, G. B. (1990). The relationship of paralinguistic cues to impression formation and the recall of medical messages. *Health Communication, 2,* 47–57. http://dx.doi.org/10.1207/s15327027hc0201_4

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology, 32,* 76–92. http://dx.doi.org/10.1177/0022022101032001009

Schroeder, J., & Epley, N. (2015). The sound of intellect: Speech reveals a thoughtful mind, increasing a job candidate's appeal. *Psychological Science, 26,* 877–891. http://dx.doi.org/10.1177/0956797615572906

Seltzer, L. J., Prososki, A. R., Ziegler, T. E., & Pollak, S. D. (2012). Instant messages vs. speech: Hormones and why we still need to hear each other. *Evolution and Human Behavior, 33,* 42–45. http://dx.doi.org/10.1016/j.evolhumbehav.2011.05.004

Singer, P. (1994). *Ethics*. Oxford, United Kingdom: Oxford University Press.

Takayama, L., & Nass, C. (2008). Driver safety and information from afar: An experimental driving simulator study of wireless vs. in-car information services. *International Journal of Human-Computer Studies, 66,* 173–184. http://dx.doi.org/10.1016/j.ijhcs.2006.06.005

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior, 33,* 210–216. http://dx.doi.org/10.1016/j.evolhumbehav.2011.09.004

Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59,* 433–460. http://dx.doi.org/10.1093/mind/LIX.236.433

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

Twenge, J. M. (2013). Does online social media lead to social connection or social disconnection? *Journal of College & Character, 14,* 11–20. http://dx.doi.org/10.1515/jcc-2013-0003

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology, 52,* 113–117. http://dx.doi.org/10.1016/j.jesp.2014.01.005

Weisbuch, M., Pauker, K., & Ambady, N. (2009). The subtle transmission of race bias via televised nonverbal behavior. *Science, 326,* 1711–1714. http://dx.doi.org/10.1126/science.1178358

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25,* 1115–1125. http://dx.doi.org/10.1177/01461672992512005

Zaki, J., Bolger, N., & Ochsner, K. (2009). Unpacking the informational bases of empathic accuracy. *Emotion, 9,* 478–487. http://dx.doi.org/10.1037/a0016551

Received May 5, 2016
Revision received June 7, 2016
Accepted July 6, 2016 ■