AI & SOCIETY
https://doi.org/10.1007/s00146-022-01398-y

ORIGINAL ARTICLE

Toward the search for the perfect blade runner: a large-scale, international assessment of a test that screens for "humanness sensitivity"

Robert Epstein[1] · **Maria Bordyug**[1,2] · **Ya-Han Chen**[1,2] · **Yijing Chen**[1,3] · **Anna Ginther**[1,4] · **Gina Kirkish**[1,5] · **Holly Stead**[1]

Received: 4 June 2021 / Accepted: 7 February 2022

Abstract
We introduce a construct called "humanness sensitivity," which we define as the ability to recognize uniquely human characteristics. To evaluate the construct, we used a "concurrent study design" to conduct an internet-based study with a convenience sample of 42,063 people from 88 countries (52.4% from the U.S. and Canada).We sought to determine to what extent people could identify subtle characteristics of human behavior, thinking, emotions, and social relationships which currently distinguish humans from non-human entities such as bots. Many people were surprisingly poor at this task, even when asked simple questions about human relationships or anatomy. Participants were best at identifying subtle aspects of human cognition and worst at identifying subtle aspects of human communication. Test scores were good predictors of whether someone was employed and modest predictors of other self-reported criterion measures. We also found that people identifying themselves in marginal societal categories (e.g., in the "other" category for gender or sexual orientation) identified themselves as less human and also scored lower on our test. As computers continue to become more human-like, our study suggests that the vast majority of humankind will likely have great difficulty distinguishing them from people. Can methods be devised for improving this ability? Might humanness sensitivity help people to make such distinctions? Will people who excel at differentiating humans and non-human entities—like the "blade runners" in the 1982 and 2017 feature films—someday hold a special place in society?

Keywords Blade runners · Humanness sensitivity · Turing test · EHI · Epstein humanness inventory

1 Introduction

In this paper, we introduce a construct called "humanness sensitivity"—the ability to recognize uniquely human characteristics—along with a test for measuring such sensitivity. This test could be considered one of the first developed for the purpose of finding people who are good at distinguishing humans from computers.

✉ Robert Epstein
re@aibrt.org

1 American Institute for Behavioral Research and Technology, 1035 E. Vista Way Ste. 120, Vista, CA 92024, USA

2 University of California San Diego, San Diego, CA, USA

3 Rice University, Houston, TX, USA

4 University of California, Berkeley, CA, USA

5 Department of Neurology, University of California, San Francisco, CA, USA

1.1 The need for new expertise

In the 1982 film, "Blade Runner," starring Harrison Ford, Ford's character, Rick Deckard, is a special kind of police officer whose job it is to identify and destroy human-like androids. Deckard is particularly good at distinguishing androids from humans, which is extremely difficult to do. The film was inspired by the 1968 book by Philip K. Dick, *Do Androids Dream of Electric Sheep*, and in both the book and the film Deckard uses a special device called the Voight–Kampff machine to help him make the determination. He ultimately falls in love with a beautiful android named Rachael, who is almost impossible to distinguish from a human—in other words, who can pass what is now sometimes called a Total Turing Test (Harnad 1989; Powers 1998; Schweizer 1998). "Blade Runner 2049," the 2017 sequel to the 1982 film, takes place 30 years after the events of the original film and revolves around what ultimately became of Rachael.

In spite of occasional claims to the contrary, we are nowhere near being able to produce a Rachael or even the clumsy human-like character, Mr. Data, from the Star Trek series (Epstein 2014, 2017; cf. McCoy 2014). We can create androids that look human (e.g., see Epstein 2006; Taylor 2016; Zlotowski, et al. 2016), that move in a human-like fashion (e.g., De Momi et al. 2016; Dillow 2012; Khatib et al. 2004; Sakai et al. 2013), and that sometimes create positive social impressions (Reeves et al. 2020). We can also create computers that can distinguish visual stimuli (e.g., Bhattacharya, and Talapatra 2005; Quiñonez et al. 2015), that emulate humor and human speech to some extent (e.g., Banks and Van Ouytsel 2020; Fukui et al. 2005; Ha 2015; Johanson et al. 2020; Kuwamura et al. 2016; Mearian 2013; Nakamura and Sawada 2006), that distinguish words spoken by humans (e.g., Bishop 2016; Hanafiah et al. 2004; Rutkin 2015; Stiefelhagen et al. 2004), and that even provide therapeutic services (e.g., David et al. 2014; Libin and Libin 2004).

The big nut, however—understanding human language—has, in our view, not yet been cracked by computers, even in the face of significant advances in natural language processing (e.g., Yan et al. 2020), computational linguistics (e.g., Boyce-Jacino and DeDeo 2021; Mitkov 2014), and machine learning (e.g., Brown et al. 2020; Marr 2020). Even though millions of humans are now communicating with bots on the internet, often without the humans knowing, the sophistication of these conversations is minimal (Hill et al. 2015). Computers that have been competing for 29 years now in the annual Loebner Prize Competition—a contest inspired by Turing's (1950) paper on "the imitation game"—have made relatively little progress (Epstein 1992, 2017; Epstein et al. 2009; Hendler and Mulvehill 2016). How and when this nut will be cracked is unclear, but even in this very difficult area, computers will only get better, not worse. As they improve, the need for experts who can distinguish computers from people will almost certainly grow.

1.2 Failing to differentiate people and computers can be problematic

Although some prominent individuals, including Stephen Hawking and Elon Musk, have said that AIs pose a grave risk to humanity (Sainato 2015), in the near future, AIs are likely to be quite helpful to humans, serving as personal assistants, guiding our vehicles, providing customer service, and performing manual labor (Gitau 2015; Meissner 2020; Murphy 2015; Muthukrishnan et al. 2017). They might even prove to perform better than humans at many tasks (Dang and Tapus 2015; West 2015). For example, an AI would presumably be able to wait indefinitely for customers to respond and remember verbatim every previous conversation it had had with specific customers (Calo 2012). An AI would also be able to access information instantly and never demand a promotion or a raise (Cui et al. 2017). Even where robots and humans prove to be equally good at the same job, robots are "more efficient, less needy and untiring in their abilities" (Danaher 2019, p. 134). These are all developments that are likely to appeal to businesses and consumers (Loebbecke and Picot 2015). In fact, the only people who might complain about this transition are the ones who will lose their jobs (Danaher 2019; Ford 2015; Smith and Anderson 2014). Whether the new AIs will be sophisticated enough to satisfy a customer's needs fully remains to be seen, but over time the AIs will almost certainly become more and more capable (Nakagawa 2015; Shukla Shubhendu and Vijay 2013).

In thinking about how humans and computers might be compared one day, Turing (1950) focused mainly on one issue, namely, whether a computer could converse with a human in way that would fool the human into thinking it was a person—an indication, Turing argued, that the computer might be conscious in a human sense—but there are many other ways in which a human might be distinguishable from a non-human. Harnad (1989, 1991, 1992) insisted that a Total Turing Test (TTT), in which the humanness of a human-like robot would be assessed would be necessary at some point and that eventually a stronger test assessing "neuromolecular indistinguishability" might need to be considered (Harnad 1992, p. 10).

While no computer has yet to pass the TTT or even the simple verbal test proposed by Turing (1950) (see Aamoth 2014; Epstein 2014; Gewirtz 2018; Knightly 2018; Neufeld and Finnestad 2020; Nieva 2018), AI has already advanced sufficiently so that people are sometimes having trouble distinguishing AIs from humans, especially online (Aron 2011; Derrick et al. 2013; Epstein 2007; West 2015). Sometimes this is because of expectations; when people are expecting and perhaps yearning to interact with a human; on dating websites, for example, they will often fail to spot a bot (Epstein 2007; Light 2016; Mansfield-Devine 2015). When the hookup website, AshleyMadison.com, was hacked in 2015, researchers found that most of the men who thought they were chatting with attractive young women were actually chatting with bots. According to journalist Annalee Newitz, "20 million men out of 31 million received bot mail, and about 11 million of them were chatted up by an automated 'engager'" (Newitz 2015). Wishful thinking can be blinding, but given that people are rarely required to think about what makes humans unique, we speculate that a lack of knowledge about human uniqueness—at least among average humans—might also contribute to the confusion.

That people are easily fooled by bots should not surprise us. The literatures on anthropomorphism and anthropocentrism remind us how strongly inclined people are to see humanity virtually everywhere, even in patterns of shadows on the moon (Cousineau 2019; Gunkel 2017; Nass et al.

2007; Preston 1991). The fact that computers and robots can sometimes mimic human intelligence or speech is bound to make us especially vulnerable.

At the moment, failing to spot a bot might lead to pointless purchases (Bilton 2014; Kelion 2014), a waste of time and resources (Kabadaian 2018; Vincent 2017), or embarrassment (Epstein 2007).[1] A witty, sassy, or funny chatbot might get people to click on a bad link or give up sensitive information (Kelion 2014; Kerr 2004). Bots also have the potential to become yet another channel for spam (Rodrigo and Abraham 2012; Stone-Gross et al. 2011). Someday—perhaps very soon, in fact—failing to spot a bot might even be dangerous (Hartzog 2014). Even now, bots are sometimes so good at collecting information from or influencing humans that they can be used for social media attacks (Huber et al. 2009; Frenkel 2017; Woolley and Gorbis 2017). The judgment a person makes about whether they are communicating with a person or a bot is important; for example, people are less likely to try to repair misunderstandings when they think they are interacting with a bot (Corti and Gillespie 2016).

Ironically, even though people might not be able to articulate what makes a person a person or a bot a bot, they are amazingly good at noticing when a human or human-like entity is *not quite right*. This implies that people's implicit knowledge about unique human characteristics might greatly exceed their explicit knowledge, the kind of dichotomy that is often recognized in dual-process theories of social cognition (Nosek 2007). People are often uncomfortable around human-like androids (Szollosy 2017; Zlotowski et al. 2016), viewing them as "eerie," particularly when there is a subtle mismatch between human and artificial features (Ho and MacDorman 2010; Mori 1970; cf. Kätsyri et al. 2015).

1.3 How are humans different?

Although we sometimes mistake bots for people, humans are in fact distinctly different from non-humans in many ways, at least at the moment. We humans are products of both our evolutionary and personal histories; computers have no such histories. We are social animals deeply affected by the cultures in which we are reared; computers do not have cultures. Our thoughts and feelings are limited by our anatomy and physiology, as well as by the capabilities of our sense organs; in theory, computers have no such limitations. Our thoughts and feelings revolve around the deeply personal relationships we have with other people; computers have no such relationships—at least so far. Humans bond to their early caregivers, and normal development might be impeded by the lack of such bonding (Harlow 1958); computers do not bond at all. We also dream and daydream; Dick's (1968) speculations about android dreams notwithstanding, computers presumably experience nothing even remotely like dreaming. Could computers somehow be programmed—or perhaps even evolve—so that they become more human-like in these ways? Perhaps (Bartneck et al. 2017; de Graaf et al. 2015), but for the time being we are truly unique in non-trivial ways (Neufeld and Finnestad 2020, p. 824).

Website security is a niche in which this issue is of constant concern. In 2000, the CAPTCHA box was invented as a way of trying to guarantee that only a human could enter a website (Von Ahn et al. 2003). The first CAPTCHA boxes used hazy or obscure text, but as AIs have become more sophisticated, CAPTCHA boxes have become increasingly complex and demanding, now sometimes requiring us to identify cars or signposts in photographs or even to drag shoes into a shoebox (Davidson et al. 2014; Hendler and Mulvehill 2016; NuCAPTCHA 2015; SolveMedia 2015; SweetCAPTCHA 2015; Vikram et al. 2011; von Ahn et al. 2003). Research has also shown that people differ in their ability to complete such tasks properly (Albert et al. 2010; Banday and Shah 2011; Belk et al. 2012; Wei et al. 2012).

That humans might have trouble spotting bots, and that some of us will do better than others at this task, should surprise no one. Think of the difficulties we have in making accurate judgments about other people—or even about *ourselves*. Humans frequently make "attribution errors"; we often blame our own misbehavior on external stimuli, for example, while blaming the misbehavior of other people on their inherent flaws or traits (Dekker 2014; Hasan and Khalid 2014; Pettigrew 1979). We also struggle for much of our lives to understand or explain our own behavior and feelings—those strong feelings of jealously our partner engenders in us, for example, or the overwhelming feelings of love we might have toward a child or, sometimes, toward a complete stranger or a pet (Imbir 2016; Smith and Lane 2016; Tryon 2014). Some of us spend years or even decades in therapy trying to gain insights into why we behave or feel as we do (Butler et al. 2006; Derlaga and Berg 2013; Kelly 2003; Rogers 2012). It should hardly surprise us to learn that we are not very insightful when it comes to knowing what makes humans unique.

1.4 Daunting programming challenges

To complicate matters further, knowing what makes humans unique—at least at a high level of expertise—must involve knowing *how unique each and every human is*. To put this another way, it is not enough to know *in general* how humans are special; one must also know how the uniqueness of each individual is likely to be expressed. If, off in

[1] Epstein was encouraged to relive his embarrassment on an episode of NPR's Radiolab, accessible here: https://www.wnycstudios.org/podcasts/radiolab/segments/137466-clever-bots.

the distance, you hear someone speaking in a high-pitched voice, you will probably surmise that you are hearing a little girl and you might even be able to estimate her age. As the voice grows nearer, so you can hear the particulars, you will quickly be able to build a picture of the child's personality and background. From a programming perspective, the challenges here are immense. A human-like AI not only must behave *like a human*, it must also behave like a *particular* human—one with a unique background and personality, with unique tastes and desires, and with unique emotions and ideas (Erebak and Turgut 2020). If it has a body, it must also have unique postural characteristics, mannerisms, expressions, and so on (Aly and Tapus 2016; Beer et al. 2017; Cassell and Bickmore 2003; Klowait 2018; Ng-Thow-Hing et al. 2010; Park et al. 2012). One of the quickest ways, perhaps, to unmask a bot might be to look for generic human features (Doran and Gokhale 2011; Turing 1950; von Ahn et al. 2003; Zabihimayvan et al. 2017).

Persuasive human analogues will need to be able to make sense of jumbled human language coming from multiple speakers (we all do this with varying degrees of success every day), must be able to use and understand ever-changing colloquialisms, and must instantly make allowances in their communications for the age, gender, ethnicity, and sensory limitations of the person with whom they are conversing (Bennett et al. 2014; Eyssel 2017; Eyssel and Hegel 2012; Eyssel and Loughnan 2013; Eyssel and Ribas 2012). Humans do all this and much more, often without being able to articulate such matters (Fusaroli et al. 2014; Gillespie and Cornish 2010; Howarth 2006; Markova 2003; Psaltis and Duveen 2006). Human analogues that are intended to fool people into thinking that they are human will also need to be able to simulate human limitations such as working memory capacity, speed of processing, motor skills limitations, and attention span limitations (Conway et al. 2002; Polderman et al. 2006; Shipstead and Broadway 2013). These abilities and limitations also vary among individuals.

1.5 Finding people who know what makes humans unique

These are complex issues, but we believe they can be managed to some extent by recognizing two things: first, that humans are indeed different from non-humans in a number of ways, and second, that people undoubtedly vary in their ability to distinguish humans from non-humans. The question of individual differences has long been raised in a number of areas of human functioning: intelligence (Engle et al. 1999; Wickens et al. 2015), personality traits (Eysenck and Eysenck 1987), cognitive processing styles (Demetriou et al. 2013; Riding and Cheema 1991), athletic ability (Mirzaei et al. 2013), and so on. In the context of Turing-type tests, no matter what the test—whether the watered down Turing Tests of the early Loebner Prize Competitions (Christian 2011; Epstein 1992; Epstein et al. 2009) or the souped-up challenge that will someday be posed in a Total Turing Test—the individuals serving as judges will presumably differ in their abilities to make the necessary discriminations. Even Turing recognized this problem, claiming that by 2000 a computer might be able to fool only an "average interrogator" (Turing 1950, p.440).

The present paper is our attempt to determine who would make a great Turing Test judge—or perhaps even a great blade runner. We are not looking at this issue comprehensively; rather, we are exploring just one aspect of it by introducing a new psychological construct we call "humanness sensitivity"—the ability to recognize uniquely human characteristics. We are also introducing a test that measures this sensitivity.

Note that although humanness sensitivity would presumably help someone to distinguish a human from a non-human, in many situations there might be easier ways to do so. To find out whether a Terminator-type android is human, one might simply use an x-ray device or see if it bleeds. As Turing (1950) anticipated, these days human-bot interactions typically occur only in the verbal realm; the physical characteristics of one's correspondent are completely hidden. This brings us back to humanness sensitivity. We believe that in the very near future—and perhaps even as you read this—the ability of a human to distinguish humans from non-humans will be important in many contexts—even essential. In the present study, we therefore measure an ability that should help someone spot a bot: the ability to recognize eight different kinds of uniquely human characteristics. Because we are able to compute separate scores for each of these eight categories, we can determine how good people are at recognizing each type of human uniqueness. We also assess some aspects of the reliability and validity of our new measurement tool, as well as delineate how humanness sensitivity varies according to the demographic characteristics of our test takers.

To determine how valuable humanness sensitivity is in everyday life, we also look at how strongly total scores on our test are correlated with a number of self-reported criterion variables, such as people's assessments of how happy and successful they are. Finally, we rank order our eight subscales according to how well they predict those same criterion variables. In so doing, we are able to determine which types of humanness sensitivity appear to have more value in everyday life.

2 Methods

2.1 Study design

Our investigation employed a "concurrent study design" that provided convergent validity evidence with related measures, following the most recent guidelines of *Standards for Educational and Psychological Testing*, co-published by the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education (American Educational Research Association 2014, p. 17). Specifically, we sought to measure the strength of the relationships between our test scores and the scores on answers to criterion questions. This design was "concurrent" because we obtained test scores and criterion measures at the same time, a strategy that avoids possible temporal confounds. Results from studies employing this design are considered especially robust when the pattern of relationships between test scores and criterion measures proves to be consistent across different demographic groups (American Educational Research Association 2014, pp. 17–18).

2.2 Test construction

Our test, called the Epstein Humanness Inventory (EHI), included 66 multiple-choice questions that asked about eight different types of characteristics that currently distinguish humans from nonhumans. The eight categories were based on a review of relevant studies in psychology and related fields; the review was initially conducted in 2011 and later expanded. The categories were as follows: (1) Interpreting Subtle Aspects of Human Culture, (2) Interpreting Subtle Emotional Experiences, (3) Interpreting Subtle Sensory Information, (4) Understanding Subtle Aspects of Human Cognition, (5) Understanding Subtle Aspects of Human Gender and Relationships, (6) Understanding Subtle Characteristics of Human Reasoning, (7) Understanding Subtle Forms of Communication, and (8) Understanding Unusual Human Physical Characteristics. Definitions, examples of test questions and relevant references are shown in Table 1. The 66 questions were presented in the same random order for all participants in the study.

2.3 Procedure

The EHI was first posted online on April 10, 2011 at https://HowHumanAreYou.com. A link to the test was also added at that time to a video showing a human-like Japanese android that had originally been posted on YouTube on April 28, 2007.[2] Over time, links to the test appeared on other websites, such as https://Reddit.com and https://Facebook.com. We had no control over where links were posted or removed.

To make the test entertaining—and, we hoped, to attract more test takers—it was written in a humorous fashion. In the opening instructions, the test was framed as follows:

> No, it's not a joke. This is a test designed to help humanity cope with a serious problem, one that is becoming more of a concern every day: On the phone, over the Internet, and even in person, *are you dealing with a human, a computer, a robot, or an alien?*
>
> And are *you* really a human, or have you been replaced by a robot, or even by an alien, without you knowing it? Has your brain been tampered with by aliens, or maybe by secret government agencies, so that you are no longer as human as you used to be? Just how human are you? That is the question….
>
> Sure, you're thinking, "No sweat!" You're as human as apple pie, right? But this is a difficult test, full of subtleties designed to ferret out the hidden truth—to separate the men from the toys, so to speak. If you're willing to put your humanness to the test, get ready to rumble. And if you don't have the stomach—assuming, that is, that you even *have* a stomach—to find out that you're not as human as you thought you were—that chemicals in your food, invisible mind control devices, or an alien abduction that you can't even remember has taken away some of your humanness, too bad! Suck it up!
>
> And if you are *not* a human, beware. You *will* fail this test, and we will find you and dissect or dismantle you, whichever seems more diabolical at the time….
>
> Before we get to the test itself, we'll ask you a few basic questions about yourself. This information is being collected to enable us to improve future versions of the test. We're also just nosey. (It's a human thing.)

We then asked demographic questions about gender, age, race, location, sexual orientation, education status and employment, after which we asked six criterion questions—that is, questions the answers to which we believed our test scores might be able to predict. Criterion questions employed 10-point Likert scales (1 = low, 10 = high). These questions asked participants to rate how much contact they generally had with other people, how much success they had had in their personal and professional lives, how happy they were, how much success they had had in their romantic relationships, and how human they considered themselves to be. Answers to such questions might be predictable from scores on the test because, at one level, the test is actually measuring *how human people are*. We are, after all, asking questions that only humans should be able to answer correctly. The higher their scores, the more easily they should be able to relate to other people and the more human they should consider themselves to be.

To preserve the anonymity of participants, no information was collected that would allow us to identify them. Because of the anonymity and the low-risk nature of the content of the survey, our study was approved as exempt under HHS

[2] The video is accessible here: https://www.youtube.com/watch?v=MY8-sJS0W1I.

Table 1 Eight categories of humanness sensitivity

1. Interpreting subtle sensory information (5 items): Identification of particular tastes; typical reactions to scary noises; reactions to pain; fears about loss of senses; and consequences of tasting/touching particular objects
Sample item:
I avoid licking ice cubes because:
a. Ice cubes are very cold
b. They remind me of someone I used to date
c. Sometimes they get mad
d. My tongue might stick to them
e. Licking them makes them melt
Correct answer: d
References: Abram (2012), Geldard et al. (1953), Haslam et al. (2008a, b), Lindsay and Norman (2013), Matthen (2015), Putnam (1994)

2. Interpreting subtle aspects of human culture (14 items): Identification of bad habits; beliefs about poetry, music and religion,; typical daily routines; social politeness; what people aim for in life; knowledge about famous characters, kings and presidents; the meaning behind behavior; places people want to go; who spends time in certain places
Sample item:
Which of these lines is very likely *not* part of a poem written by a human?
a. Bursting upward, outward, in jagged lines, the word rises to the light
b. My favorite pizza toppings are pepperoni and mushrooms
c. Do intentions count when there's no one there to know them?
d. Love me, and the world will shift in a loving direction
e. I replied like a pinball, bouncing post to post
Correct answer: b
References: Cassirer (1972), Cortes et al. (2005), Haslam et al. (2008a, b), Lindblom and Ziemke (2003), Nass and Moon (2000), Pande (1965), Provine (2001)

3. Interpreting subtle emotional experiences (8 items): Knowledge about situation-specific emotions; emotions related to colors; how one feels when experiencing particular emotion; recognizing how people feel or behave when experiencing particular emotions
Sample item:
If a woman came out of a department store feeling very, very guilty, that probably means:
a. She had just gotten fired
b. She had just had an argument with her boyfriend
c. She had just had her purse stolen
d. She had probably stolen something from the store
e. She was probably Catholic
Correct answer: d
References: Brave and Nass (2003), Demoulin et al. (2004), Dolan (2002), Haslam et al. (2008a, b), Haslam et al. (2008a, b), Leyens et al. (2001), Wierzbicka (1986), Withers and Vernon (2006)

4. Understanding subtle characteristics of human reasoning (7 items): Knowledge about objects or food that fit together; how people think; how people make sense of complex situations
Sample item:
How might a bright young woman who is pretty good with numbers reply if she were asked to multiply 203 times 598 in her head?
a. I think the answer is about 120,000
b. The answer is exactly 121,394
c. I think the answer is about 140
d. Please go away
e. I think the answer is about a billion
Correct answer: a
References: Anderson (2017), Brewka et al. (1997), Evans (2003), Gathercole (2003), Gigerenzer and Goldstein (1996), Holyoak and Morrison (2005), Johnson-Laird (2010), Oaksford and Chater (2007), Reason (1990)

5. Understanding subtle aspects of human cognition (4 items): Knowing how words are related to each other; recognizing similarities between letter shapes and famous monuments; being able to remember one's responses on earlier questions; knowing about the content of human dreams
Sample item:
Which of the following is least likely for someone to experience in a dream?
a. Flying without the aid of an airplane
b. Seeing a tiger transform into a fish
c. Chatting with a dead loved one
d. Multiplying large numbers
e. Growing an extra arm
Correct answer: d
References: Loughnan and Haslam (2007), Missakabo (1998), Moll et al. (2005), Pio-Abreu et al. (2015), Premack (2010), Taylor and Brown (1988), Tomasello and Rakoczy (2003)

Table 1 (continued)

6. Understanding subtle aspects of human gender and relationships (10 items): Knowing about women's menstrual cycles; knowing about male/male competition; knowing about males and females typically dress; knowing how men seduce women; know about human dating; knowing about biological differences between genders; knowing about friendships and romance

Sample item:

When Jim learned that John had recently purchased a 200 HP lawnmower, Jim responded in a manly way—that is, by:

a. Purchasing a 201 HP lawnmower
b. Purchasing a 300 HP lawnmower
c. Poking his eye out with a fork
d. Purchasing a 100 HP lawnmower
e. Purchasing a .50 caliber machine gun

Correct answer: b

References: Brownlie (2006), Costrich et al. (1975), Eagly and Wood (1999), Heilman (1979), Lemaster et al. (2017), Tannen (1990), Vaes and Paladino (2010), Wood (1997)

7. Understanding subtle forms of communication (10 items): Knowing what forms of communication are acceptable; knowing likely responses to people based on their demographic characteristics; knowing which responses are typical or likely in different situations; knowing how people respond to annoying people; knowing how parents talk to children or how lovers talk to each other

Sample item:

What is a sweet, gentle, nurturing mom most likely to say to her little 3-month-old baby?

a. Oh, you're so cute
b. Would you like the left one or the right one?
c. That's certainly a cute baby you have
d. Grow up
e. Ooh you little cutie pootie wootie

Correct answer: e

References: Beattie and Ellis (2017), Carey (2008), Ekman (1993), Hauser et al. (2002), Lasswell (1948), Littlejohn and Foss (2010), Ting-Toomey and Chung (2005), Tomasello (2010), Trenholm (2017)

8. Understanding unusual human physical characteristics (8 items): Knowing how human anatomy limits our ability to carry or lift things; knowing how different parts of the body can be used; knowing how many people it will take to lift heavy objects; knowing which body parts are most valuable; knowing how the body reacts to strong emotions; knowing what distracts people

Sample item:

If you saw a graphic news report about a mass murder, what part of your body might start to feel uncomfortable?

a. Kidney
b. Back
c. Eyes
d. Hands
e. Stomach

Correct answer: e

References: Friedman (2010), Gaby (2008), Karnath et al. (2005), Schachter and Singer (1962), Young (2001)

regulations by the federally registered Institutional Review Board of our hosting institution (the American Institute for Behavioral Research and Technology).

Following the demographic and criterion questions, participants were told, "For each of the following questions, select what you believe is the best possible answer a human can give. If you are truly human, that should n't be too hard. If you are less than human, or something other than human, you're screwed." The 66 multiple-choice questions followed (see Table 1 for sample questions, or visit the full test at https://HowHumanAreYou.com). Upon completing these questions, participants could get their score by clicking a SUBMIT button.

In keeping with the light-hearted nature of the test, the results showed participants where their score—presented as percent correct—put them on a scale from "alien" at the low extreme to "über human" at the high extreme (Fig. 1). The score was followed by a serious debriefing about the potential scientific value of the test results (see "Debriefing" in the Appendix).

2.4 Sample

Our sample consisted of 42,063 people from 88 countries (52.4% from the U.S. and Canada) who took our online test between April 17, 2011 and April 23, 2020. We had no control over who took our test, and hence this should be considered a convenience sample (see Discussion for our comments on the advantages and disadvantages of having such a sample).

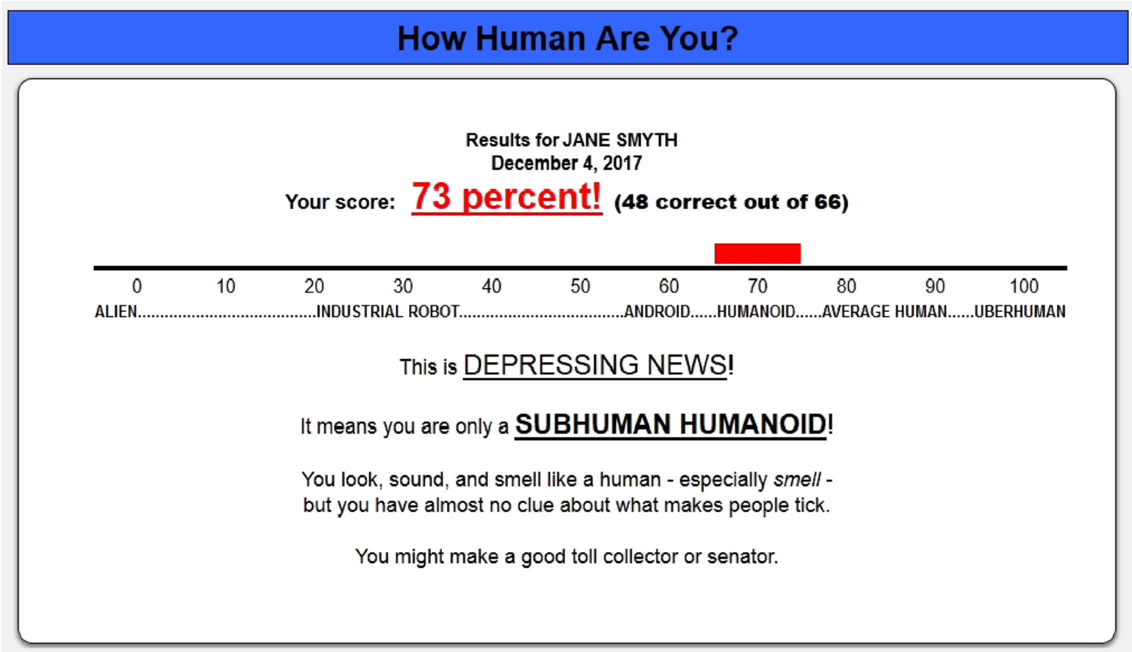

Fig. 1 Screenshot of the type of result given after completion of the test

Before beginning our analysis, we removed 16,021 cases from our database as a result of a cleaning process. Cases were removed because (a) less then half of the test questions were answered, (b) English fluency was less than 6 on a scale from 1 to 10 (where 10 indicated highest fluency), or (c) the age that was given was under 12. Participants under age 12 ($n=395$) were not included because our questions had a Flesch-Kincaid grade level of 6.0. In the U.S., students generally complete grade 6 when they are 12, so most 12-year-olds (at least in the U.S.) should be able to comprehend the EHI. (We acknowledge that many people never learn to read properly and that people of any age might have trouble comprehending our test.) We also eliminated cases when people took the test more than once in one day, retaining only the first instance in which they answered more than half the questions.

22,024 (52.4%) of our participants were from the US and Canada, 17,582 (41.8%) were from other countries, and 2457 (5.8%) did not provide their locations. The mean age was 22.4 (SD = 10.7). 24,976 (59.4%) of our participants identified themselves as male, 13,067 (31.1%) as female, 1453 (3.5%) as other, 1298 (3.1%) as unsure, and 1268 (3.0%) did not provide their gender. 29,649 (70.5%) of our participants identified themselves as white, 4152 (9.9%) as Asian, 2232 (5.3%) as Hispanic, 1101 (2.6%) as Black, 329 (0.8%) as American Indian, 3470 (8.2%) as other, and 1130 (2.7%) did not specify their race. 25,203 (59.9%) of our participants identified themselves as straight, 5858 (13.9%) as bisexual, 2338 (5.6%) as gay, 2864 (6.8%) as other, 3757 (8.9%) as unsure, and 2043 (4.9%) did not identify their sexual orientation.

24,830 (59.0%) of our participants said they were not employed, 14,422 (34.3%) said they were employed, and 2811 (6.7%) did not provide information about their employment. 9336 (22.2%) of our participants said they had not completed high school, 18,537 (44.1%) said they had completed high school, 3909 (9.3%) said they had received an associates-level degree, 6879 (16.4%) said they had completed college, 2217 (5.3%) said they had a masters degree, 741 (1.8%) said they had a doctorate, and 444 (1.0%) did not provide information about their education.

3 Results

3.1 Reliability and validity evidence

Because our test was conducted online and preserved the anonymity of test takers, we could not measure test–retest reliability. We also did not develop an alternative form of the test, so alternative-form reliability could not be estimated. However, internal-consistency reliability was fairly high, as indicated by both Cronbach's alpha (0.87) and the Guttman split-half test (0.85).

Evidence that our test was indeed measuring the construct "humanness sensitivity" was suggested by how well test scores were correlated with the answers people gave to six criterion questions—questions about happiness, success,

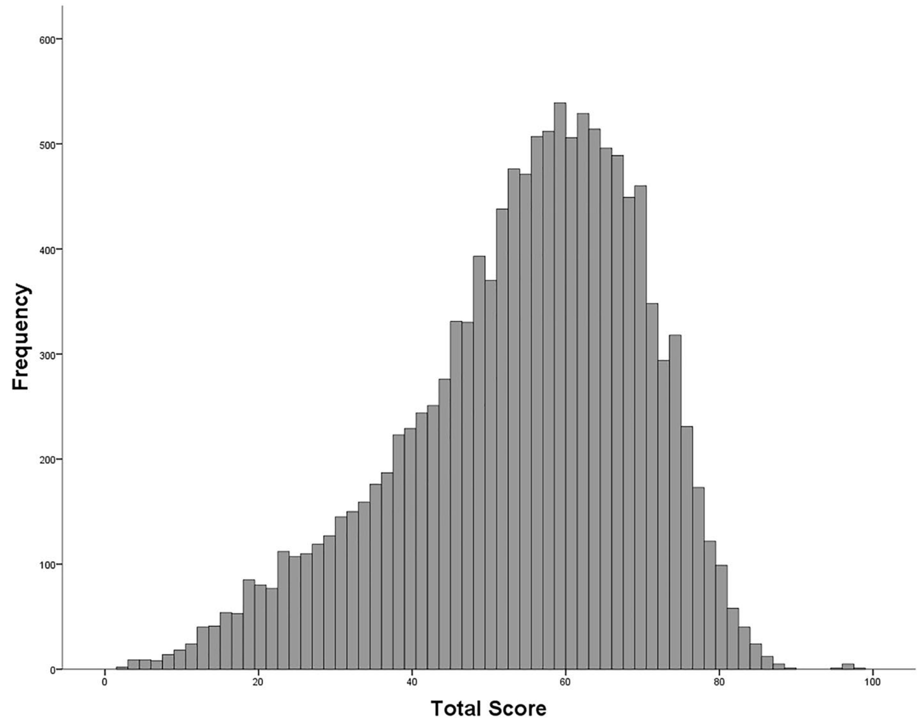

Fig. 2 A histogram showing how individual total scores were distributed

Table 2 Mean test scores for each of the eight humanness sensitivity categories

Category	Mean (SD)
Understanding subtle aspects of human cognition	70.1 (24)
Interpreting subtle emotional experiences	62.2 (23.1)
Interpreting subtle aspects of human culture	61.1 (17.9)
Understanding unusual human physical characteristics	58.9 (23)
Interpreting subtle sensory information	54.1 (26)
Understanding subtle characteristics of human reasoning	52.5 (21.5)
Understanding subtle aspects of human gender and relationships	47.7 (18.6)
Understanding subtle forms of communication	46.7 (19.8)

and so on. Test scores were moderately correlated with the self-reported level of humanness (Spearman's $\rho = 0.32$, $P < 0.001$).[3] The correlations were lower for five other criterion questions, but all of them were positive and highly significant: self-reported frequency of contact with other people: $\rho = 0.16$, $P < 0.001$; self-reported success in romantic relationships: $\rho = 0.11$, $P < 0.001$; self-reported success in one's personal life: $\rho = 0.10$, $P < 0.001$; self-reported success in one's professional life: $\rho = 0.08$, $P < 0.001$; and self-reported happiness: $\rho = 0.09$, $P < 0.001$. In addition, test scores were positively correlated with age ($\rho = 0.17$, $P < 0.001$), which suggests that people become more knowledgeable about human uniqueness as they grow older.

Although we had no a priori reason to believe that test scores would predict employment status, it is notable that people who were employed outscored people who were not: $M_{yes} = 58.1$ (SD = 15.3), $M_{no} = 54.8$ (SD = 15.2), Mann–Whitney $U = 15.43 * 10^7$, $P < 0.001$, Cohen's $d = 0.22$.

3.2 Competency means

The average total score on the test was 55.9% (SD = 15.4), and a histogram of individual scores was negatively skewed (median = 59) (Fig. 2). The average scores for each of the

[3] Because scores on the EHI are on an ordinal scale of measurement, nonparametric statistics, such as Spearman's ρ, the Mann–Whitney U, and the Kruskal–Wallis H, were used throughout this report. Means and standard deviations are reported for comparison purposes, although the appropriateness of their use with ordinal data has long been debated (e.g., Lord 1953; Townsend and Ashby 1984). Test scores are always given as a percentage of the maximum possible raw score.

eight categories of humanness sensitivity we measured are shown in Table 2. Participants scored highest on Understanding Subtle Aspects of Human Cognition (70.1%) and lowest on Understanding Subtle Forms of Human Communication (46.7%).

3.3 Demographic differences

3.3.1 Gender and sexual orientation

For both gender and sexual orientation, one notable finding was that people who identified themselves as being outside mainstream categories rated themselves on average to be less human than people in those categories; perhaps even more surprising, they also generally scored lower on the test (see Discussion for further details about this issue). We found an overall effect for gender, for example ($M_{male}=57.0$ [SD = 14.7], $M_{female}=55.4$ [SD = 15.2], $M_{other}=50.6$ [SD = 18.9], $M_{unsure}=46.4$ [SD = 19.2], Kruskal–Wallis $H=566$, $P<0.001$), but the difference between scores for males and females was relatively small (1.6% points) ($M_{male}=57.0$ [SD = 14.7], $M_{female}=55.4$ [SD = 15.2], $U=15.23 * 10^7$, $P<0.001$, $d=0.11$); whereas, the difference between scores for males and females combined versus scores for those who labeled themselves "other" or "unsure" combined was substantially larger (7.8% points) ($M_{male/female}=56.4$[SD = 14.9], $M_{other/unsure}=48.6$ [SD = 19.1], $U=4.02 * 10^7$, $P<0.001$, $d=0.46$). Similarly, we found an overall gender effect in response to the question, "How human are you?" ($M_{male}=7.2$ [SD = 2.7], $M_{female}=6.8$ [SD = 2.8], $M_{other}=5.4$ [SD = 3.1], $M_{unsure}=5.1$ [SD = 3.1], $H=1012.9$, $P<0.001$), but the difference between scores for males and females was relatively small (0.4 points): $M_{male}=7.2$ [SD = 2.7], $M_{female}=6.8$ [SD = 2.8], $U=15.04 * 10^7$, $P<0.001$, $d=0.14$; whereas, the difference between scores for males and females combined versus scores for those who labeled themselves "other" or "unsure" combined was substantially larger (1.8 points): ($M_{male/female}=7.0$ [SD = 2.8], $M_{other/unsure}=5.2$ [SD = 3.1], $U=3.52 * 10^7$, $P<0.001$, $d=0.61$.

We found the same pattern for sexual orientation. Although there was an overall effect ($M_{straight}=37.5$ [SD = 9.7], $M_{gay}=36.7$ [SD = 10.5], $M_{bisexual}=37.4$ [SD = 10.0], $M_{other}=34.5$ [SD = 11.5], $M_{unsure}=34.6$ [SD = 11.0], $H=342.4$, $P<0.001$), there was only a marginal effect for gays versus straights ($M_{straight}=37.5$ [SD = 9.7], $M_{gay}=36.7$ [SD = 10.5], $U=2.86 * 10^7$, $P=0.027$, $d=0.07$); whereas, there was a substantially larger effect for combined scores for gays, straights, and bisexuals versus combined scores for those who labeled themselves "other" or "unsure" ($M_{straight/gay/bi}=56.7$ [SD = 14.9], $M_{other/unsure}=52.4$ [SD = 17.0], $U=9.48 * 10^7$, $P<0.001$, $d=0.27$). Similarly, there was an overall effect for sexual orientation in response to the question, "How human are you?" ($M_{straight}=7.2$ [SD = 2.7], $M_{gay}=6.7$ [SD = 2.9], $M_{bisexual}=6.7$ [SD = 2.8], $M_{other}=5.7$ [SD = 3.0], $M_{unsure}=6.1$ [SD = 3.0], $H=1169.2$, $P<0.001$), but the gay/straight difference for this measure ($M_{straight}=7.2$ [SD = 2.7], $M_{gay}=6.7$ [SD = 2.9], $U=2.67 * 10^7$, $P<0.01$, $d=0.17$) was somewhat smaller than the difference between scores for gays, straights, and bisexuals combined versus those who labeled themselves "other" or "unsure" combined ($M_{straight/gay/bi}=7.1$ [SD = 2.8], $M_{other/unsure}=5.9$ [SD = 3.0], $U=2.67 * 10^7$, $P<0.001$, $d=0.42$).

3.3.2 Race, country, and education

Effects were found for race both on the total score measure ($M_{White}=57.3$ [SD = 15.1], $M_{Black}=52.9$ [SD = 14.8], $M_{Hispanic}=53.1$ [SD = 14.7], $M_{Asian}=53.5$ [SD = 14.5], $M_{AmerIndian}=50.5$ [SD = 15.0], $M_{Other}=50.5$ [SD = 17.6], $H=961.9$, $P<0.001$) and the self-reported level of humanness measure ($M_{White}=7.0$ [SD = 2.8], $M_{Black}=6.5$ [SD = 3.0], $M_{Hispanic}=6.9$ [SD = 2.8], $M_{Asian}=6.9$ [SD = 2.8], $M_{AmerIndian}=6.2$ [SD = 3.0], $M_{Other}=6.0$ [SD = 3.2], $H=328.5$, $P<0.001$). We also found an effect for country ($M_{US/Canada}=57.7$ [SD = 15.3], $M_{Other}=54.0$ [SD = 15.1], $U=18.63 * 10^7$, $P<0.001$, $d=0.24$).

An effect was also found for level of education completed. Generally speaking, the more education people had, the higher they scored on our test ($M_{none}=53.3$ [SD = 16.0], $M_{highschool}=55.8$ [SD = 14.9], $M_{associates}=56.7$ [SD = 14.6], $M_{college}=59.0$ [SD = 14.8], $M_{masters}=59.0$ [SD = 14.9], $M_{doctorate}=56.1$ [SD = 19.9]; $\rho=0.111$, $P<0.001$; $H=621.3$, $P<0.001$) and the more highly they rated their own level of humanness ($M_{none}=6.7$ [SD = 3.0], $M_{highschool}=6.9$ [SD = 2.9], $M_{associates}=6.7$ [SD = 2.9], $M_{college}=7.3$ [SD = 2.6], $M_{masters}=7.4$ [SD = 2.6], $M_{doctorate}=7.0$ [SD = 3.2]; $\rho=0.052$, $P<0.001$; $H=202.0$, $P<0.001$).

3.4 Change in test scores over time

Because our data were collected over a period of 9 years, we looked for trends in the 8 calendar years for which we had full years of data (2012 to 2019). The mean total score in 2012 was 58.5%, whereas the mean total score in 2019 was 53.0%. However, there was no consistent decrease in scores over the years. Table 3 shows the changes in the number of test takers each year, along with the changes in the mean total scores, ages, fluency levels, percentage of whites (the largest racial group in the study), and percentage of males (the largest gender group in the study). Because of the inconsistent variability in these variables, we evaluated the change in test scores over the years by dividing our sample into two groups—one prior to October 21st, 2015 (the midpoint in time between the first and last days in our sample), and

Table 3 Changes by year (for years in which a full year of data were available) in number of test takers, mean total scores, age, fluency levels, percentage of whites and percentages of males

Year	*n*	Mean total score (SD)	Mean age (SD)	Mean fluency level (SD)	Percentage of whites	Percentage of males
2012	370	58.5 (14.6)	26.7 (13.4)	9.6 (1.0)	75.4	48.4
2013	324	57.7 (16.0)	26.4 (14.6)	9.6 (0.9)	70.4	41.4
2014	553	55.9 (15.2)	22.2 (11.4)	9.5 (0.9)	72.9	41.6
2015	1866	54.8 (15.3)	20.6 (8.9)	9.4 (1.0)	68.6	36.9
2016	5882	53.7 (15.6)	22.5 (9.9)	9.4 (1.0)	66.8	42.8
2017	3917	52.4 (15.9)	23.5 (12.5)	9.4 (1.0)	64.7	41.1
2018	23,437	57.6 (14.8)	21.9 (9.8)	9.1 (1.1)	72.9	72.4
2019	3809	53.0 (16.2)	23.3 (12.0)	9.2 (1.1)	67.6	45.2

Table 4 Results of six linear regressions showing which categories of humanness sensitivity best predict self-reported criterion variables

Criterion variable (self-reported)	Best predictor variable among the eight categories of humanness sensitivity[a]	Standardized beta	t	Adjusted r^2
Humanness	Interpreting subtle emotional experiences	0.317***	68.66	0.101
Contact with others	Interpreting subtle emotional experiences	0.161***	32.97	0.026
Relationship success	Understanding unusual human physical characteristics	0.092***	18.98	0.008
Professional success	Understanding unusual human physical characteristics	0.076***	15.58	0.006
Happiness	Understanding subtle forms of communication	0.086***	17.64	0.007
Personal success	Understanding subtle forms of communication	0.093***	19.08	0.009

***$P < 0.001$

[a]Produced in the single-component model of a stepwise regression

one from October 21st, 2015 to April 23, 2020, the last day in our sample. We then computed the mean total score for each of those groups. The difference between the means was marginally significant, but it was small in absolute terms (0.6%), and the effect size was small ($M_1 = 56.5$ [SD = 15.3], $M_2 = 55.9$ [SD = 15.4], $U = 6.63 * 10^7$, $P < 0.05$, $d = 0.04$). Although our year-to-year samples and scores varied considerably, it appears that test scores, on average, were reasonably stable over the period we examined.

3.5 Regressions used to predict criterion variables

Stepwise linear regressions were used to find which of the eight categories of humanness sensitivity best predicted each of our six criterion variables (Table 4). Both the self-reported level of humanness (standardized $\beta = 0.317$, $P < 0.001$) and the self-reported frequency of human contact ($\beta = 0.161$, $P < 0.001$) were best predicted by our Interpreting Subtle Emotional Experiences scale. Understanding Unusual Human Physical Characteristics proved to be the best predictor of two of the criterion variables: success in romantic relationships ($\beta = 0.092$, $P < 0.001$) and professional success ($\beta = 0.076$, $P < 0.001$), although the r^2 values were small (Table 4). Happiness was best predicted by our Understanding Subtle Forms of Human Communication scale ($\beta = 0.086$, $P < 0.001$), and so was personal success ($\beta = 0.093$, $P < 0.001$); again, the r^2 values were small.

4 Discussion

The current study opens the door to a new type of measurement that will likely become increasingly important in the near future: identifying characteristics of people that might help them to distinguish humans from computers (in whatever forms computers might eventually take). We have begun this task by focusing on one particular type of skillset we call humanness sensitivity, or the ability to recognize uniquely human characteristics.

Before we examine some of the more interesting outcomes of this study, we will first look at its limitations. The eight categories of unique human characteristics we incorporated into our test—(1) Interpreting Subtle Aspects of Human Culture, (2) Interpreting Subtle Emotional Experiences, (3) Interpreting Subtle Sensory Information, (4) Understanding Subtle Aspects of Human Cognition, (5) Understanding Subtle Aspects of Human Gender and Relationships, (6) Understanding Subtle Characteristics of Human Reasoning, (7) Understanding Subtle Forms of Communication, and (8) Understanding Unusual Human Physical Characteristics—are somewhat overlapping and not definitive. For example, at the moment, humans are unique in being composed entirely of organic matter. As writer Stanislaw Lem reminded us in his classic novel, *The Cyberiad*, humans arose from "noxious exhalations," "putrid excrescences" and "creeping molds that slithered forth from the ocean onto land" (Lem 1974, pp. 283–284). They lived "by devouring one another" until, eventually, they "stood upright supporting their globby substance by means of calcareous scaffolding" (p. 284). Yet the EHI contains no category for "Recognizing that Humans Are Organic." In spite of its light tone, the EHI also lacks a category one might label, "Recognizing a Human's Ability to Appreciate Wit and Humor" (or, to put it more simply, "Understanding Jokes"). At the moment, computers are notoriously bad at emulating or appreciating human wit (Binsted et al. 2006; Hernandez 2016; Kao et al. 2015; Niculescu et al. 2013; Ritchie 2001; Tay et al. 2016; cf. Johanson et al. 2020). Needless to say, our categories and items were also designed for English-speaking Westerners; they might have looked very different had we designed them for people in non-Western cultures (American Psychology Association 2017).

Besides the limitations of our categories, our study was also limited by the nature of its sample: a convenience sample of people who found our test online and then chose to complete it. The changes in the mean total scores and demographic characteristics from year to year (Table 3) are reminders of one of the limitations of conducting long-term studies on the internet: We had no control over the makeup of our sample. Its changing composition over time was probably due to changes in the various websites that linked to the test, as well as to changes in how search engines ranked the test; we had no control over such factors.

On the upside, the shifting links and search rankings undoubtedly gave us a more diverse sample of test takers—42,063 people from 88 countries. There is also increasing evidence that people who take anonymous tests online generally give more honest responses than they do when their identities are known or when a human test administrator is present, especially when they are being asked about socially sensitive topics (Dillman et al. 2014; Durant et al. 2002; Dwight and Fiegelson 2000; Joinson 1999; Kaplan and Saccuzzo 2009; Krumpal 2013; Ong and Weiss 2000; Robertson et al. 2017). We submit that although our sample is not ideal, it is almost certainly more representative of humanity than is the proverbial group of 200 sophomores in the subject pool of a single university (Anderson et al. 2012; Goodwinand Goodwin 2018, pp. 144–146; Henrich et al. 2010; cf. Kühberger et al. 2014).

Regarding our results, one finding was particularly robust and also disturbing—namely, that people who identified themselves as being in minority categories of gender ("other" or "unsure") or sexual orientation ("other" or "unsure") not only rated themselves as being less human than people in mainstream categories, they also scored lower on the EHI. The low scores people gave themselves on our self-reported humanness scale can perhaps be explained simply: People in what society sometimes regards as unacceptable categories of gender or sexual orientation often struggle with low self-esteem, identity issues, or mental health problems (Birkett et al. 2009; Eisenberg et al. 2017; Joel et al. 2014; Richards et al. 2016; Zhao et al. 2010)—problems likely produced, at least in part, by pressures to conform to societal norms (Blakemore 2003; Roberts et al. 2013; Smith and Juvonen 2017; Toomey et al. 2012). But why did people in these categories also score lower on the EHI? In other words, why would doubts about one's humanity make one less aware of what people dream about, how people reason, or how people communicate? Our study was not designed to address such issues, and we are reluctant to speculate. However, given the magnitude of the effects we found for both gender ($M_{\text{male/female}} = 56.4$ [SD = 14.9], $M_{\text{other/unsure}} = 48.6$ [SD = 19.1], $U = 4.02 * 10^7$, $P < 0.001$, $d = 0.46$) and sexual orientation ($M_{\text{straight/gay/bi}} = 56.7$ [SD = 14.9], $M_{\text{other/unsure}} = 52.4$ [SD = 17.0], $U = 9.48 * 10^7$, $P < 0.001$, $d = 0.27$), we believe this is an issue that deserves further study.

How meaningful are our eight categories of humanness sensitivity? One way we explored this issue was to use linear regressions to see which sensitivity categories best predicted each of six criterion variables (Table 4). Self-reported level of humanness was best predicted by Interpreting Subtle Emotional Experiences ($\beta = 0.32$, $P < 0.001$); since only humans can experience human emotions, this makes sense. Interpreting Subtle Emotional Experiences was also the best predictor of the frequency of contact our test takers reported having with other people ($\beta = 0.16$, $P < 0.001$); that, too, is reasonable, but one might think that Understanding Subtle Forms of Communication would be at least as helpful. The latter sensitivity was the best predictor of level of personal success, and that too seems reasonable. The other three criterion variables—success in romantic relationships, professional success, and happiness—were best predicted by Understanding Unusual Human Physical Characteristics. While some knowledge of human anatomy might be helpful in advancing a romantic relationship, we offer no explanation for why this category of sensitivity might contribute to happiness or professional success.

Future versions of the EHI might use a different scoring strategy. We chose to use the "one-correct-response" or "best-response" scoring strategy that is commonly used to score major U.S. standardized tests such as the Scholastic Aptitude Test (SAT) (Korsunsky undated-a, undated-b). The best-response scoring strategy is popular in part because it tends to produce reliable test scores (Carneson et al. 2016; Korsunsky undated-a, undated-b; Sadler 1998; Towns 2014; Treagust 1988). As is common with multiple-choice tests, many of our items included both a "best response" and one or more "distractors"—that is, answers that are fairly reasonable—maybe even correct—but that are not the best response (Carneson et al. 2016; Kehoe, 1995; Korsunsky undated-a, undated-b; Sadler 1998). Consider the following item from Table 1, for example:

> How might a bright young woman who is pretty good with numbers reply if she were asked to multiply 203 times 598 in her head?
> a. I think the answer is about 120,000.
> b. The answer is exactly 121,394.
> c. I think the answer is about 140.
> d. Please go away.
> e. I think the answer is about a billion.

The best answer is *a*, although *b* and *d* are also pretty good. Answer *b* ("121,394") is, in fact, the correct answer to the multiplication problem; it is inferior to *a* ("about 120,000") only because a computer can perform the calculation easily, whereas a human can do so only with some difficulty. Answer *d* ("Please go away") is, on the other hand, a sassy, human thing to say. Answer *a* is only marginally better than *d* because it makes a point that *d* does not (and hence is more informative) —namely, that a human could fairly easily *estimate* the correct answer even if he or she could not easily compute the correct answer.

Our point here is that the best-answer scoring strategy does not extract all the information we might be able to obtain about our test takers. In theory, we could extract more nuanced information about them by giving partial credit for certain responses, or even by computing subscores from specific human-typical answers such as "Please go away." When one is using multiple-choice tests to compute someone's course grade or SAT score, nuanced information of this sort might serve no purpose, but when one is trying to find competent blade runners, a deeper understanding of the test takers might be helpful, if not essential. The test instrument used in the present study could not be used to explore the merits of a partial-credit approach, because the test items were not designed with partial credit in mind. Future research could explore such possibilities.

One might also wonder whether the light tone of the test content could have had a systematic impact on participants' scores. If people were taking the test just to have fun, is it possible that they were not performing as well as they might have given more serious content? Our design gives us no way to answer that question, but we also have no reason to believe that some of the more interesting findings in this study—the fact that people scored highest in the category labeled Understanding Subtle Aspects of Human Cognition, for example, or the fact that people in minority categories of gender or sexual orientation scored significantly lower on the test—were artifacts of the light tone of the content. If the light tone of the content had any systematic impact on the results, it is not clear what that is.

Given the high level of skill one would ultimately want in a blade runner, any paper-and-pencil type of test such as a multiple-choice or Likert-scale test could only be used for screening purposes. Ultimately, one would subject high scorers to demanding simulations in which they had to correctly differentiate humans from bots, AIs, or robots. Over time, presumably, as intelligent machines become more human-like, these simulations will become more demanding, perhaps requiring people to use specialized software or hardware (such as the fictional Voight–Kampff machine) to assist them in their evaluations.

We also acknowledge that in a real search for competent blade runners, one would undoubtedly want to start out with far more demographic information than we collected, for example, information about people's professions, educational backgrounds, and experience with computers and artificial intelligence. With detailed demographic information, test scores could be used to make inferences about which kinds of backgrounds shape or attract competent blade runners.

How valuable is humanness sensitivity? The positive correlations we found between our test scores and our six criterion measures suggest that awareness of what makes humans unique has value in everyday life—perhaps, as we speculated earlier, because humanness sensitivity allows people to relate better to other people.

Overall, people scored fairly poorly on our test ($M = 55.9\%$), and the range of scores was extreme (from 0 to 100%). These findings suggest that, absent training programs that might improve humanness sensitivity, finding a good blade runner might be difficult. Will tests such as the EHI be used someday to identify the next Rick Deckard? And can humanness sensitivity—or perhaps a more general set of skills that allow us to distinguish humans from bots, AIs, and androids—be trained? What would such training entail? Again, the present study sheds no light, but the EHI or similar instruments could certainly be used as part of training programs to measure improvements in relevant abilities.

To determine whether humanness sensitivity might have value in a Turing Test competition, one might (a) administer the EHI, (b) have test takers serve as judges in a Turing Test contest, and then (c) determine whether EHI test scores are correlated with measures of how well the judges performed in the contest. This sort of procedure is a logical next step to take in assessing the value and validity of the EHI. Thinking ahead, it might also be interesting at some point to see how well bots or AIs can score on the EHI. Thinking even further ahead, a time may come when AIs will outperform humans in distinguishing humans from computers, just as they can now outperform radiologists in evaluating X-rays (Rajpurkar et al. 2017).

Is humanness sensitivity correlated with other, more standard, measures of human ability—with emotional intelligence or some aspects of general intelligence, for example? Such correlations undoubtedly exist, but we speculate that competent blade runners or Turing Test judges will need to have such highly specialized skills—skills that will undoubtedly have to evolve over time to parallel advances in computing—that specialized methods of measuring such skills will continue to have value for the foreseeable future. At first glance, the skills needed to distinguish humans from non-human intelligent entities might seem analogous to the specialized skills people in law enforcement use to detect lies (Holm 2010; Mares and Turvey 2018; Vrij et al. 2017), but the analogy quickly breaks down when one realizes how difficult the bot problem really is. The signals for detecting human lies do not change much over time, but the non-humanness of AIs and bots is necessarily a moving target. Just as SEO techniques constantly evolve to outfox search engines, so too will intelligent systems evolve to outfox humans, which is why Deckard's job was so difficult. Recall Harnard's (1992) suggestion that someday "neuromolecular indistinguishability" will need to be evaluated when we are trying to unmask nonhuman intelligence. At some point, even that criterion might fail. If, someday, humans and androids prove to be indistinguishable at the molecular level, will it even be meaningful to say that androids still exist as a separate class of beings? If, over time, humans are gradually replaced by such beings, would anyone even notice?

Returning to one of the central questions we raised earlier in this essay, what would programmers have to do to construct a truly humanlike computer or android? And why, for that matter, would a programmer or tech company go to the trouble of doing so? After all, to pass the TTT, a computer would have to have a massively complex history—entirely fabricated—a distinct personality, a quirky sense of humor, aches and pains, a faulty memory, digestion problems, highly limited senses, flawed reasoning abilities, more than a few neuroses, occasional suicidal thoughts, a reasonable dose of selfishness, and strong desires to masturbate and copulate. Why create such a monstrosity? Why, in particular, would one tamper with one of the computer's greatest strengths—its precise and nearly infallible memory?

Although it may seem odd, there is good reason to create such machines, and it can even be argued that humans have been dreaming of such machines for thousands of years (Zarkadakis 2016). One of the most profitable ways in which computers are being used these days is in simulating young women who just cannot wait to have sex with human men who have credit cards (Epstein 2007; Light 2016; Mansfield-Devine 2015). The more realistic the simulation, the faster the money will flow. Thinking ahead, there will also be a large market for general-purpose humanlike androids to serve as companions or servants, the more humanlike the better. The most valuable androids will *not* sink us into the uncomfortable pit of the uncanny valley; rather, they will simulate human quirks, failings and flaws extremely well, just like the neurotic C-3PO robot does in the Star Wars movies. Bear in mind that computers that simulate human behavior well will be increasingly difficult to distinguish from people, thus increasing the demand for competent blade runners.

How urgent is the need for tests like the EHI? To put this another way, how quickly are AIs, bots, and robots becoming part of our lives, both in ways that might benefit us and ways that might hurt us? In the original "Blade Runner" story, androids (called "replicants") are banned from use on earth because they are autonomous and have superhuman strength. They are banned even though they are designed to have life spans of only 4 years—a failsafe hardwired into replicants to protect humanity. In other depictions of AI-dominated futures AIs are depicted as serious threats, for example, in the disturbing 1966 film, "Colossus: The Forbin Project," in which a sentient computer uses nuclear blackmail to conquer humanity, or in the 2004 film "I, Robot," in which robots take over humanity to save humanity from itself, or even in the classic film "2001: A Space Odyssey," in which the computer HAL 9000 becomes insane and murders most of the crew of the spaceship *Discovery One*. How close are we to such scenarios?

The truth is that it's hard to say. Today's AIs are often easy to unmask (Boyle 2016; Coniam 2008; Hill et al. 2015; Jankel 2015; Yu 2017), sometimes pathetically so (Epstein 2017; Kontos 2016; Zerega 2017), but advances in machine learning and other new technologies are moving so rapidly that anything is possible. In some respects, advanced conversational programs such as Steve Worswick's Mitsuku (winner of the 2013, 2016 and 2017 Loebner Prize Competitions) (Kontos 2016; Yu 2017), Apple's Siri (Bellegarda 2014), and Amazon's Alexa (Fischer et al. 2016) are still using the same simple programming tricks that Joseph Weizenbaum's Eliza used back in the 1960s (Weizenbaum 1966); they are far from being conscious, and in no sense do they truly understand human language. But it is only a matter of time before these lines are crossed. In the meantime, we are confident that blade-runner-type professionals will begin to surface, along with evaluative tests for identifying people who excel at such work.

Appendix

Debriefing

What this test is all about (seriously): nonhumans are likely to have great difficulty answering questions about unique human characteristics: our informality, idiosyncrasies, and individual styles, for example. Even more difficult for a nonhuman to fathom: extremely subtle aspects of human relationships and emotions, as well as how these and other human phenomena change as we get older. Humans also make predictable errors; when computer programs are written that imitate people, they always incorporate a serious dose of "artificial stupidity"—spelling, arithmetic, and reasoning errors, for example. To be human is to err.

Humans have bizarre dreams and daydreams. We have food cravings, especially when pregnant. We laugh at funny jokes but also when someone slips on a banana peel (what's funny about that?). Music, art, literature and even sports sometimes make us giddy. Our memories change over time; properly designed computer memories do not. Many of us are propelled through life in a quest for money, sex, or power, or, in some cases, the perfect cup of coffee. When deeply in love, we are sometimes completely insane. We seek happiness, but some of us remain deeply depressed for months or years no matter what we do, sort of like Marvin the robot in *The Hitchhiker's Guide to the Galaxy*—except that we're real.

Many of us stubbornly believe in God or the supernatural, no matter what the facts. We sometimes become needy or whiny when sick or injured, and we feel profoundly embarrassed if we fart at the wrong time. We tell lies—but only when it's "absolutely convenient," as the British comedian Benny Hill put it. We divide the world into good and evil forces—both of which see themselves as good—and we sometimes commit crimes. We get headaches and tummy aches, and our hearts sometimes race when we spot an old lover. Some of us, sometimes, get tipsy or even drunk, and many of us deliberately alter our usual states of consciousness with just about any drug we can get our hands on.

It is difficult to imagine an alien, robot, or computer being able to answer any but the most trivial questions about such matters. To answer the tough questions about humans, one needs to be one.

Even with that advantage, however, most people score well under 100% on this test, mainly because, among our other foibles, not all of us understand the nuances of human relationships, emotions, defects, and idiosyncrasies. In that sense—if humanness sensitivity can be defined as "the ability to recognize uniquely human characteristics"—some of us are more human than others. There is good news, however: for the time being, even the least human of us is still more human than the most human computer.

This test was created by Dr. Robert Epstein, a distinguished research psychologist, the founder of the Cambridge Center for Behavioral Studies, and the co-creator and first director of the Loebner Prize Competition in Artificial Intelligence, an annual Turing Test in which judges try to distinguish humans from computers, first held at The Computer Museum in Boston, Massachusetts in 1991. You can learn more about the Turing Test in Dr. Epstein's book, *Parsing the Turing Test: Methodological and Philosophical Issues in the Quest for the Thinking Computer*. You can download his article, "My Date with a Robot" (from *Scientific American Mind*) here: https://DrRobertEpstein.com/pdf/Epstein-My_Date_With_a_Robot-Scientific_American_Mnd-2006.pdf. To view Dr. Epstein interacting with a beautiful Japanese android, click here: https://www.youtube.com/watch?v=MY8-sJS0W1I

Test results are being used in an ongoing study of "humanness sensitivity" being conducted by the American Institute for Behavioral Research and Technology (AIBRT) in Vista, California, USA. AIBRT is a nonpartisan, nonprofit, 501(c)(3) research organization dedicated to improving human life. The study was approved by AIBRT's federally registered Institutional Review Board. If you have any questions about the study or would like to be informed about the outcome of the study after the results are made public, please contact us at tests@aibrt.org. Include "EHI study" in your subject line.

Acknowledgements This report is based in part on data presented at the 97th annual meeting of the Western Psychological Association, Sacramento, CA, April 2017. Epstein and Kirkish originally described the concept of humanness sensitivity, along with pilot data, in a paper presented at the 42nd annual meeting of the Society for Computers in Psychology, Minneapolis, MN, November 2012 (Epstein and Kirkish, 2012). We thank Paul McKinney and Ronald E. Robertson for assistance in preparing an early version of the EHI. We also thank Bruce Edmonds, James Hendler and George Zarkadakis for helpful feedback on an earlier version of the manuscript.

Funding This research was supported by general funds of the American Institute for Behavioral Research and Technology, a nonprofit, nonpartisan, 501(c)(3) organization. The research was not supported by any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Availability of data and material Data requests should be sent to info@aibrt.org. The full data set has not been posted publically to protect the anonymity of the study participants, a requirement of the exempt status granted by the sponsoring institution's IRB.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflicts of interest to report.

References

Aamoth D (2014) Interview with Eugene Goostman, the fake kid who passed the Turing Test. TIME, June 9. https://time.com/2847900/eugene-goostman-turing-test/

Abram D (2012) The spell of the sensuous: perception and language in a more-than-human world. Vintage Books, USA

Albert D, Jeng B, Tseng C, Wang J (2010) A study of CAPTCHA and its application to user authentication. Proc Int Conf Comput Collective Intell 6422:433–440

Aly A, Tapus A (2016) Towards an intelligent system for generating an adapted verbal and nonverbal combined behavior in human–robot interaction. Auton Robot 40(2):193–209

American Educational Research Association (2014) Standards for educational and psychological testing. American Educational Research Association, American Psychological Association, National Council on Measurement in Education

American Psychological Association (2017) Cultural differences. Particularly exciting experiments in psychology. https://www.apa.org/pubs/highlights/peeps/issue-96

Anderson J, Burks S, Carpenter J, Gotte L, Maurer K, Nosenzo D, Potter R, Rocha K, Rustichini A (2012) Self-selection and variations in the laboratory measurement of other-regarding preferences across subject pools: evidence from one college student and two adult samples. Exp Econ 16:170–189. https://doi.org/10.1007/s10683-012-9327-7

Aron J (2011) Software tricks people into thinking it is human. New Scientist, September 6. https://www.newscientist.com/article/dn20865-software-tricks-people-into-thinking-it-is-human

Banday MT, Shah NA (2011) Challenges of CAPTCHA in the accessibility of Indian regional websites. In: Proceedings of the fourth annual ACM Bangalore conference, vol 31. https://doi.org/10.1145/1980422.1980453

Banks J, Van Ouytsel J (2020) Cybersex with human- and machine-cued partners: gratifications, shortcomings, and tensions. Technol Mind Behav 1(1)

Bartneck C, Lyons MJ, Saerbeck M (2017) The relationship between emotion models and artificial intelligence. arXiv. https://arxiv.org/pdf/1706.09554.pdf

Beattie G, Ellis AW (2017) The psychology of language and communication. Routledge, New York

Beer JM, Liles KR, Wu X, Pakala S (2017) Affective human–robot interaction. In: Emotions and affect in human factors and human-computer interaction, pp 359–381. https://doi.org/10.1016/B978-0-12-801851-4.00015-X

Belk M, Fidas C, Germanakos P, Samaras G (2012) Do cognitive styles of users affect preference and performance related to CAPTCHA challenges? In: CHI'12 extended abstracts on human factors in computing systems. ACM, pp 1487–1492. https://doi.org/10.1145/2212776.2223660

Bellegarda JR (2014) Spoken language understanding for natural interaction: The Siri experience. In: Mariani J, Rosset S, Garnier-Rizet M, Devillers L (eds) Natural interaction with robots, knowbots and smartphones. Springer, Switzerland

Bennett CC, Sabanovic S, Fraune MR, Shaw K (2014) Context congruency and robotic facial expressions: do effects on human perceptions vary across culture? In: Proceedings of the 23rd IEEE international symposium on robot and human interactive communication, pp 465–470. https://doi.org/10.1109/ROMAN.2014.692629

Bhattacharya S, Talapatra S (2005) Robot motion planning using neural networks: a modified theory. Int J Lateral Comput 2(1):9–13

Bilton N (2014) Social media bots offer phony friends and real profit. New York Times, November 19. https://www.nytimes.com/2014/11/20/fashion/social-media-bots-offerphony-friends-and-real-profit.html

Binsted K, Bergen B, O'Mara D, Coulson S, Nijholt A, Stock O, Strapparava C, Ritchie G, Manurung R, Pain H, Waller A, O'Mara D (2006) Computational humor. IEEE Intell Syst 21(2):59–69

Birkett M, Espelage DL, Koenig B (2009) LGB and questioning students in schools: the moderating effects of homophobic bullying and school climate on negative outcomes. J Youth Adolesc 38:989–1000

Bishop R (2016). 2016 World Robot Conference: freakishly human-like robots understand speech and can read facial expressions. Daily Mirror, October 24. https://www.mirror.co.uk/news/world-news/2016-world-robot-conference-freakishly-9115782

Blakemore JEO (2003) Children's beliefs about violating gender norms: boys shouldn't look like girls, and girls shouldn't act like boys. Sex Roles 48(9):411–419

Boyce-Jacino C, DeDeo S (2021) Cooperation, interaction, search: computational approaches to the psychology of asking and answering questions. PsyArXiv. https://psyarxiv.com/5mgn2/download

Boyle D (2016). What's the difference between robots and humans? It's my newt. The Guardian, September 22. https://www.theguardian.com/commentisfree/2016/sep/22/robots-pass-humans-turing-test-loebner-prize

Brave S, Nass C (2003) Emotion in human–computer interaction. Hum Comput Interact 4:53–68

Brewka G, Dix J, Konolige K (1997) Nonmonotonic reasoning: an overview. CSLI Publications, Stanford

Brown T, Mann B, Ryder N, Subbiah M, Kaplan J, Dhariwal P, Neelakantan A, Shyam P, Sastry G, Askell A, Agarwal S, Herbert-Voss A, Krüger G, Henighan T, Child R, Ramesh A, Ziegler D, Wu J, Winter C, Hesse C, Chen M, Sigler E, Litwin M, Gray S, Chess B, Clark J, Berner C, McCandlish S, Radford A, Sutskever I, Amodei D (2020) Language models are few-shot learners. ArXiv. https://arxiv.org/abs/2005.14165

Brownlie E (2006). Adults' constructions of gender: a lifespan Q methodological study. ProQuest. https://search.proquest.com/docview/304943064?accountid=11862

Butler AC, Chapman JE, Forman EM, Beck AT (2006) The empirical status of cognitive-behavioral therapy: a review of meta-analyses. Clin Psychol Rev 26(1):17–31

Calo MR (2012) Robots and privacy. In: Lin P, Abney K, Bekey GA (eds) Robot ethics: the ethical and social implications of robotics. MIT Press, Cambridge, pp 187–202

Carey JW (2008) Communication as culture, revised edition: essays on media and society. Routledge, New York

Carneson J, Delpierre G, Masters K (2016) Designing and managing multiple choice questions, 2nd edn. Routledge, London

Cassell J, Bickmore T (2003) Negotiated collusion: modeling social language and its relationship effects in intelligent agents. User Model User-Adap Inter 13:89–132

Cassirer E (1972) An essay on man: An introduction to a philosophy of human culture. Yale University Press, New Haven

Christian B (2011) The most human human: What talking with computers teaches us about what it means to be alive. Doubleday, New York

Coniam D (2008) Evaluating the language resources of chatbots for their potential in English as a second language. ReCALL 20(1):98–116

Conway AR, Cowan N, Bunting MF, Therriault DJ, Minkoff SR (2002) A latent variable analysis of working memory capacity, short-term memory capacity, processing speed, and general fluid intelligence. Intelligence 30(2):163–183

Cortes BP, Demoulin S, Rodriguez RT, Rodriguez AP, Leyens JP (2005) Infrahumanization or familiarity? Attribution of uniquely human emotions to the self, the ingroup, and the outgroup. Pers Soc Psychol Bull 31(2):243–253

Corti K, Gillespie A (2016) Co-constructing intersubjectivity with artificial conversational agents: people are more likely to initiate repairs of misunderstandings with agents represented as human. Comput Hum Behav 58:431–442

Costrich N, Feinstein J, Kidder L, Maracek J, Pascale L (1975) When stereotypes hurt: three studies of penalties in sex-role reversals. J Exp Soc Psychol 11:520–530

Cousineau MJ (2019) Limiting the discourse of computer and robot anthropomorphism in a research group. AI Soc 34:877–888. https://doi.org/10.1007/s00146-018-0836-6

Cui L, Huang S, Wei F, Tan C, Duan C, Zhou M (2017) Super agent: a customer service chatbot for e-commerce websites. In: Proceedings of ACL 2017, system demonstrations, pp 97–102. https://doi.org/10.18653/v1/P17-40176

Danaher J (2019) The rise of the robots and the crisis of moral patiency. AI Soc 34:129–136. https://doi.org/10.1007/s00146-017-0773-9

Dang T, Tapus A (2015) Stress game: the role of motivational robotic assistance in reducing user's task stress. Int J Soc Robot 7(2):227–240

David D, Matu S-A, David OA (2014) Robot-based psychotherapy: concepts development, state of the art, and new directions. Int J Cogn Ther 7(2):192–210. https://doi.org/10.1521/ijct.2014.7.2.192

Davidson D, Renaud K, Li S (2014) jCAPTCHA: accessible human validation. In: Proceedings of the international conference on computers helping people with special needs. Springer-Verlag, pp 129–136

de Graaf MM, Allouch SB, van Dijk JAGM (2015) What makes robots social? A user's perspective on characteristics for social human-robot interaction. In: International conference on social robotics. Springer, pp 184–193. https://doi.org/10.1007/978-3-319-2554-5_19

De Momi E, Kranendonk L, Valenti M, Enayati N, Ferrigno G (2016) A neural network-based approach for trajectory planning in robot–human handover tasks. Front Robot AI 3:34. https://doi.org/10.3389/frobt.2016.00034

Dekker S (2014) The field guide to understanding "human error." Ashgate Publishing, Farnham

Demetriou A, Spanoudis G, Shayer M (2013) Developmental intelligence: from empirical to hidden constructs. Intelligence 41(5):744–749

Demoulin S, Leyens JP, Paladino MP, Rodriguez RT, Rodriguez AP, Dovidio JF (2004) Dimensions of "uniquely" and "non-uniquely" human emotions. Cogn Emot 18:71–96

Derlaga VJ, Berg JH (eds) (2013) Self-disclosure: theory, research, and therapy. Springer Science and Business Media, New York

Derrick DC, Meservy TO, Jenkins JL, Burgoon JK, Nunamaker Jr JF (2013). Detecting deceptive chat-based communication using typing behavior and message cues. ACM Trans Manage Inf Syst 4(2):Article No. 9. https://doi.org/10.1145/2499962.2499967

Dick PK (1968) Do androids dream of electric sheep? Doubleday, New York

Dillman DA, Smyth JD, Christian LM (2014) Internet, phone, mail, and mixed-mode surveys: the tailored design method. John Wiley and Sons, New Jersey

Dillow C (2012) Video: a robot with a human skeleton. Popular Science. http://www.popsci.com/technology/article/2012-04/video-can-worlds-most-human-robotic-body-help-crack-code-human-robot-intelligence

Dolan RJ (2002) Emotion, cognition, and behavior. Science 298(5596):1191–1194

Doran D, Gokhale SS (2011) Web robot detection techniques: overview and limitations. Data Min Knowl Disc 22:183–210. https://doi.org/10.1007/s10618-010-0180-z

Durant LE, Carey MP, Schroder KEE (2002) Effects of anonymity, gender, and erotophilia on the quality of data obtained from self-reports of socially sensitive behaviors. J Behav Med 25(5):438–467

Dwight SA, Fiegelson ME (2000) A quantitative review of the effect of computerized testing on the measurement of social desirability. Educ Psychol Measur 60:340–360

Eagly AH, Wood W (1999) The origins of sex differences in human behavior: evolved dispositions versus social roles. Am Psychol 54(6):408

Eisenberg ME, Gower AL, McMorris BJ, Rider GN, Shea G, Coleman E (2017) Risk and protective factors in the lives of transgender/gender nonconforming adolescents. J Adolesc Health 61(4):521–526

Ekman P (1993) Facial expression and emotion. Am Psychol 48(4):384

Engle RW, Kane MJ, Tuholski SW (1999) Individual differences in working memory capacity and what they tell us about controlled attention, general fluid intelligence, and functions of the prefrontal cortex. In: Miyake A, Shah P (eds) Mechanisms of active maintenance and executive control. Cambridge University Press, Cambridge

Epstein R (1992) The quest for the thinking computer. AI Mag 13(2):80–95

Epstein R (2006) My date with a robot. Scientific American Mind, June/July, pp 68–73. https://drrobertepstein.com/pdf/Epstein-My_Date_With_a_Robot-Scientific_American_Mnd-2006.pdf

Epstein R (2007) From Russia, with love: how I got fooled (and somewhat humiliated) by a computer. Scientific American Mind, October/November, pp 16–17. https://drrobertepstein.com/downloads/FROM_RUSSIA_WITH_LOVE-Epstein-Sci_Am_Mind-Oct-Nov2007.pdf

Epstein R (2014). Claims that the Turing test has been passed are nonsense [Letter]. The Guardian, June 11. https://www.theguardian.com/technology/2014/jun/11/turing-test-computer-simulates-boy

Epstein R (2017) Google's fighting hate and trolls with a dangerously mindless AI. Fast Company, August 31. https://www.fastcompany.com/40459339/google-perspective-fighting-hate-and-trolls-with-a-mindless-a-i

Epstein R, Kirkish G (2012) How good are humans at distinguishing humans from computers? In: Paper presented at the 42nd annual meeting of the Society for Computers in Psychology, November, Minneapolis, MN. https://aibrt.org/downloads/EPSTEIN_&_KIRKISH_2012-Distinguishing_humans_from_computers-Soc_for_Computers_in_Psych.pdf

Epstein R, Roberts G, Beber G (eds) (2009) Parsing the Turing test: methodological and philosophical issues in the quest for the thinking computer. Springer, New York

Erebak S, Turgut T (2020) The mediator role of robot anxiety on the relationship between social anxiety and the attitude toward interaction with robots. AI Soc 35(4):1047–1053

Evans JSB (2003) In two minds: dual-process accounts of reasoning. Trends Cogn Sci 7(10):454–459

Eysenck HJ, Eysenck MW (1987) Personality and individual differences. Plenum, New York

Eyssel F (2017) An experimental psychological perspective on social robotics. Robot Auton Syst 87:363–371. https://doi.org/10.1016/j.robot.2016.08.029

Eyssel F, Hegel F (2012) (S)he's got the look: gender-stereotyping of social robots. J Appl Soc Psychol 42:2213–2230. https://doi.org/10.1111/j.1559-1816.2012.00937.x

Eyssel F, Loughnan S (2013) "It don't matter if you're Black or White?" Effects of robot appearance and user prejudice on evaluations of a newly developed robot companion. Lect Notes Comput Sci 8239:442–431. https://doi.org/10.1007/978-3-319-02675-6_42

Eyssel F, Ribas X (2012) How to be good (or bad): on the fakeability of dehumanization and prejudice against outgroups. Group Process Intergroup Relat 15:804–812. https://doi.org/10.1177/1368430212447204

Fischer M, Menon S, Khatib O (2016) From bot to bot: using a chat bot to synthesize robot motion. AAAI Fall Symposia. https://web.stanford.edu/~smenon/professional_files/publications/2016-MF-SM-OK-AAAI-HRI.pdf

Ford M (2015) Rise of the robots: technology and the threat of a jobless future. Basic Books, New York

Frenkel S (2017). Hackers hide cyberattacks in social media posts. The New York Times, May 28. https://www.nytimes.com/2017/05/28/technology/hackers-hide-cyberattacks-in-social-media-posts.html

Friedman BH (2010) Feelings and the body: the Jamesian perspective on autonomic specificity of emotion. Biol Psychol 84(3):383–393

Fukui K, Nishikawa K, Ikeo S, Shintaku E, Takada K, Takanobu H, Honda M, Takanishi A (2005) Development of a talking robot with vocal cords and lips having human-like biological structures. In: Paper presented at the 2005 IEEE/RSJ international conference on intelligent robots and systems, August, Edmonton, Alta., Canada. https://doi.org/10.1109/IROS.2005.1545356

Fusaroli R, Gangopadhyay N, Tylén K (2014) The dialogically extended mind: language as skillful intersubjective engagement. Cogn Syst Res 29:31–39

Gaby A (2008) Gut feelings: locating intellect, emotion and life force in the Thaayorre body. Culture, body, and language: conceptualizations of internal body organs across cultures and languages, pp 28–44

Gathercole S (2003) Short-term and working memory. Taylor and Francis, Hove

Geldard FA, O'Hehir R, Gavens D (1953) The human senses. Wiley, New York

Gewirtz D (2018) Google Duplex beat the Turing test: Are we doomed? ZDNet, May 14. https://www.zdnet.com/article/google-duplex-beat-the-turing-test-are-we-doomed/

Gigerenzer G, Goldstein DG (1996) Reasoning the fast and frugal way: models of bounded rationality. Psychol Rev 103(4):650

Gillespie A, Cornish F (2010) Intersubjectivity: towards a dialogical analysis. J Theory Soc Behav 40(1):19–46

Gitau B (2015) Smart hotel: Japan opens a hotel run by robots. Christian Science Monitor, July 18. https://www.csmonitor.com/Technology/2015/0718/Smart-hotel-Japan-opens-a-hotel-run-by-robots

Goodwin K, Goodwin J (2018) Research in psychology: methods and design, 8th edn. John Wiley and Sons, New York

Gunkel DJ (2017) The machine question: critical perspectives on Ai, robots, and ethics. MIT Press

Ha A (2015) Meet Musio, a robot that talks back. Tech Crunch, May 31. https://techcrunch.com/2015/05/31/meet-musio/

Hanafiah ZM, Yamazaki C, Nakamura A, Kuno Y (2004) Human-robot speech interface understanding inexplicit utterances using vision. In: Paper presented at the CHI 2004 extended abstracts on human factors in computing systems, Vienna, Austria, April. https://doi.org/10.1145/985921.986054

Harlow HF (1958) The nature of love. Am Psychol 13(12):673

Harnad S (1989) Minds, machines and Searle. J Exp Theor Artif Intell 1(1):5–25. https://doi.org/10.1080/09528138908953691

Harnad S (1991) Other bodies, other minds: a machine incarnation of an old philosophical problem. Mind Mach 1(1):43–54. https://doi.org/10.1007/BF00360578

Harnad S (1992) The Turing test is not a trick: Turing indistinguishability is a scientific criterion. SIGART Bull 3(4):9–10. https://doi.org/10.1145/141420.141422

Hartzog W (2014) Unfair and deceptive robots. Maryland Law Rev 74:785

Hasan SS, Khalid R (2014) Academic locus of control of high and low achieving students. J Res Reflect Educ 8(1):22–23

Haslam N, Kashima Y, Loughnan S, Shi J, Suitner C (2008a) Subhuman, inhuman, and superhuman: contrasting humans with nonhumans in three cultures. Soc Cogn 26(2):248–258

Haslam N, Loughnan S, Kashima Y, Bain P (2008b) Attributing and denying humanness to others. Eur Rev Soc Psychol 19(1):55–85

Hauser MD, Chomsky N, Fitch WT (2002) The faculty of language: what is it, who has it, and how did it evolve? Science 298(5598):1569–1579

Heilman ME (1979) High school students' occupational interest as a function of projected sex ratios in male-dominated occupations. J Appl Psychol 64:275–279

Hendler J, Mulvehill AM (2016) Social machines: the coming collision of artificial intelligence, social networking, and humanity. Apress, California

Henrich J, Heine S, Norenzayan A (2010) The weirdest people in the world? Behav Brain Sci 33(2–3):61–83. https://doi.org/10.1017/S0140525X0999152X

Hernandez D (2016) It's hard work being funny—especially for robots. Splinter News, January 1. https://splinternews.com/its-hard-work-being-funny-especially-for-robots-1793853950

Hill J, Ford WR, Farreras IG (2015) Real conversations with artificial intelligence: a comparison between human–human online conversations and human–chatbot conversations. Comput Hum Behav 49:245–250. https://doi.org/10.1016/j.chb.2015.02.026

Ho C, MacDorman KF (2010) Revisiting the uncanny valley theory: developing and validating an alternative to the Godspeed indices. Comput Hum Behav 26(6):1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Holm HJ (2010) Truth and lie detection in bluffing. J Econ Behav Organ 76(2):318–324

Holyoak KJ, Morrison RG (eds) (2005) The Cambridge handbook of thinking and reasoning. Cambridge University Press, England

Howarth C (2006) A social representation is not a quiet thing: exploring the critical potential of social representations theory. Br J Soc Psychol 45(1):65–86

Huber M, Kowalskiy S, Nohlbergz M, Tjoa S (2009) 2009 International conference on computational science and engineering. Vancouver, BC, pp 117–124. https://doi.org/10.1109/CSE.2009.205

Imbir KK (2016) From heart to mind and back again. A duality of emotion overview on emotion-cognition interactions. New Ideas Psychol 43:39–49

Jankel NS (2015) AI vs. Human intelligence: why computers will never create disruptive innovations. Huffington Post, April 26. https://www.huffingtonpost.com/nick-seneca-jankel/ai-vs-human-intelligence-_b_6741814.html

Joel D, Tarrasch R, Berman Z et al (2014) Queering gender: studying gender identity in 'normative' individuals. Psychol Sex 5:291–321

Johanson DL, Ahn HS, Lim J, Lee C, Sebaratnam G, MacDonald BA, Broadbent E (2020) Use of humor by a healthcare robot positively affects user perceptions and behavior. Technol Mind Behav. https://doi.org/10.1037/tmb0000021

Johnson-Laird PN (2010) Mental models and human reasoning. Proc Natl Acad Sci 107(43):18243–18250

Joinson A (1999) Social desirability, anonymity, and internet-based questionnaires. Behav Res Methods Instrum Comput 31(3):433–438

Kabadaian H (2018) How bots steal your online advertising budget. Entrepreneur. https://www.entrepreneur.com/article/313943

Kao JT, Levy R, Goodman ND (2015) A computational model of linguistic humor in puns. Cogn Sci 40:1270–1285

Kaplan RM, Saccuzzo DP (2009) Psychological testing: principles, applications, and issues. Wadsworth, Belmont

Karnath HO, Baier B, Nägele T (2005) Awareness of the functioning of one's own limbs mediated by the insular cortex? J Neurosci 25(31):7134–7138
Kätsyri J, Förger K, Mäkäräinen M, Takala T (2015) A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. Front Psychol. https://doi.org/10.3389/fpsyg.2015.00390
Kehoe J (1995) Writing multiple-choice test items. ERIC/AE Digest. Eric.ed.gov. https://eric.ed.gov/?id=ED398236
Kelion L (2014) Tinder accounts spammed by bots masquerading as singles. BBC, April 2. https://www.bbc.com/news/26850761
Kelly G (2003) The psychology of personal constructs, volume two: clinical diagnosis and psychotherapy. Routledge, New York
Kerr I (2004) Bots, babes, and Californication of commerce. Univ Ottawa Law Technol J 1:285
Khatib O, Sentis L, Park J, Warren J (2004) Whole-body dynamic behavior and control of human-like robots. Int J Humanoid Rob 1(1):29–43. https://doi.org/10.1142/S0219843604000058
Klowait N (2018) The quest for appropriate models of human-likeness: anthropomorphism in media equation research. AI Soc 33:527–536. https://doi.org/10.1007/s00146-017-0746-z
Knightly W (2018) Google Duplex: Does it pass the Turing Test? HBS Dig Init, November 13. https://digital.hbs.edu/platform-rctom/submission/google-duplex-does-it-pass-the-turing-test
Kontos J (2016) Machine consciousness and question answering. SCIREA J Comput Sci Technol 1(1). https://article.scirea.org/pdf/52006.pdf
Korsunsky B (undated-a) Multiple-choice tests: why the controversy? College Board. https://apcentral.collegeboard.org/series/multiple-choice-tests/why-controversy
Korsunsky B (undated-b) MCTs: good for assessment, good for research, good for teaching. College Board. https://apcentral.collegeboard.org/series/multiple-choice-tests/good-assessment-good-research-good-teaching
Krumpal I (2013) Determinants of social desirability bias in sensitive surveys: a literature review. Qual Quant 47(4):2025–2047
Kühberger A, Fritz A, Scherndl T (2014) Publication bias in psychology: a diagnosis based on the correlation between effect size and sample size. PLoS One 9(9). https://doi.org/10.1371/journal.pone.0105825
Kuwamura K, Nishio S, Sato S (2016) Can we talk through a robot as if face-to-face? Long-term fieldwork using teleoperated robot for seniors with Alzheimer's disease. Front Psychol 7:1066. https://doi.org/10.3389/fpsyg.2016.01066
Lasswell HD (1948) The structure and function of communication in society. Commun Ideas 37:215–228
Lem S (1974) The cyberiad. Seabury Press, New York
Lemaster P, Delaney R, Strough J (2017) Crossover, degendering, or…? A multidimensional approach to life-span gender development. Sex Roles 76(11–12):669–681
Leyens JP, Rodriguez-Perez A, Rodriguez-Torres R, Gaunt R, Paladino MP, Vaes J (2001) Psychological essentialism and the differential attribution of uniquely human emotions to ingroups and outgroups. Eur J Soc Psychol 31:395–411
Libin AV, Libin EV (2004) Person-robot interactions from the robopsychologists' point of view: the robotic psychology and robotherapy approach. Proc IEEE 92(11):1789–1803. https://doi.org/10.1109/JPROC.2004.835366
Light B (2016) The rise of speculative devices: Hooking up with the bots of Ashley Madison. First Monday 21:6. https://doi.org/10.5210/fm.v21i6.6426
Lindblom J, Ziemke T (2003) Social situatedness of natural and artificial intelligence: Vygotsky and beyond. Adapt Behav 11(2):79–96
Lindsay PH, Norman DA (2013) Human information processing: an introduction to psychology. Academic Press, New York
Littlejohn SW, Foss KA (2010) Theories of human communication. Waveland Press, USA
Loebbecke C, Picot A (2015) Reflections on societal and business model transformation arising from digitization and big data analytics: a research agenda. J Strateg Inf Syst 24(3):149–157
Lord FM (1953) On the statistical treatment of football numbers. Am Psychol 8:750–751
Loughnan S, Haslam N (2007) Animals and androids: implicit associations between social categories and nonhumans. Psychol Sci 18:116–121
Mansfield-Devine S (2015) The Ashley Madison affair. Netw Secur 9:8–16. https://doi.org/10.1016/S1353-4858(15)30080-5
Mares AC, Turvey BE (2018) The psychology of lying. False allegations: Investigative and forensic issues in fraudulent reports of crime. Academic Press, Cambridge, pp 21–36
Markova I (2003) Dialogicality and social representations: the dynamics of mind. Cambridge University Press, England
Marr B (2020) What is GPT-3, and why is it revolutionizing artificial intelligence? Forbes, Oct 5. https://www.forbes.com/sites/bernardmarr/2020/10/05/what-is-gpt-3-and-why-is-it-revolutionizing-artificial-intelligence/
Matthen M (ed) (2015) The Oxford handbook of philosophy of perception. Oxford Handbooks
McCoy T (2014) A computer just passed the Turing test in landmark trial. The Washington Post, June 9. https://www.washingtonpost.com/news/morning-mix/wp/2014/06/09/a-computer-just-passed-the-turing-test-in-landmark-trial/?utm_term=.439359416147
Mearian L (2013). NAO robot can now speak 19 languages—thanks to the cloud. Computer World, October 30. https://www.computerworld.com/article/2485467/emerging-technology/nao-robot-can-now-speak-19-languages-thanks-to-the-cloud.html
Meissner G (2020) Artificial intelligence: consciousness and conscience. AI Soc 35:225–235. https://doi.org/10.1007/s00146-019-00880-4
Mirzaei A, Nikbakhsh R, Sharififar F (2013) The relationship between personality traits and sport performance. Eur J Exp Biol 3(3):439–442
Missakabo MB (1998) Can computers do what humans do? A comparison between artificial intelligence and human intelligence. ProQuest. https://search.proquest.com/docview/304476986?accountid=11862
Mitkov R (ed) (2014) The Oxford handbook of computational linguistics, 2nd edn. Oxford University Press, Oxford
Moll J, Zahn R, de Oliveira-Souza R, Krueger F, Grafman J (2005) The neural basis of human moral cognition. Nat Rev Neurosci 6(10):799–809
Mori M (1970) The uncanny valley. Energy 7(4):33–35
Murphy M (2015) Amazon tests out robots that might one day replace warehouse workers. Quartz, July 1. https://qz.com/416668/amazon-tests-out-robots-that-might-one-day-replace-warehouse-workers/
Muthukrishnan M, Tomkins A, Heck L, Geramifard A, Agarwal D (2017) The future of artificially intelligent assistants. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, August, pp 33–34. https://doi.org/10.1145/3097983.3120999
Nakagawa D (2015) David McAfee: the second machine age is approaching. Here's how we can prepare. Huffington Post, February 24. https://www.huffingtonpost.com/dawn-nakagawa/andrew-mcafee-machine-age_b_6743660.html
Nakamura M, Sawada H (2006) Talking robot and the analysis of autonomous voice acquisition. Paper presented at the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China. https://doi.org/10.1109/IROS.2006.282257

Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56(1):81–103

Nass CI, Lombard M, Henriksen L, Steuer J (2007) Anthropocentrism and computers. Behav Inf Technol 14(4):229–238. https://doi.org/10.1080/01449299508914636

Neufeld E, Finnestad S (2020) In defense of the Turing test. AI Soc 35(2):819–827. https://doi.org/10.1007/s00146-020-00946-8

Newitz A (2015). Ashley Madison code shows more women, and more bots. Gizmodo, August 31. https://gizmodo.com/ashley-madison-code-shows-more-women-and-more-bots-1727613924

Ng-Thow-Hing V, Luo P, Okita S (2010) Synchronized gesture and speech production for humanoid robots. In: Proceedings of the IEEE/RSJ international conference on intelligent robots and systems (IROS), Taipei, Taiwan. https://doi.org/10.1109/IROS.2010.5654322

Niculescu A, van Dijk B, Nijholt A, Li H, See SL (2013) Making social robots more attractive: the effects of voice pitch, humor and empathy. Int J Soc Robot 5(2):171–191

Nieva R (2018) Alphabet chairman says Google Duplex passes Turing test in one specific way. CNET, May 10. https://www.cnet.com/news/alphabet-chairman-says-google-duplex-passes-turing-test-in-one-specific-way-io-2018/

Nosek B (2007) Implicit-explicit relations. Curr Dir Psychol Sci 16(2):65–69. https://doi.org/10.1111/j.1467-8721.2007.00477.x

NuCAPTCHA (2015) NuCAPTCHA—adaptive CAPTCHA authentication. NuCAPTCHA. https://www.nucaptcha.com

Oaksford M, Chater N (2007) Bayesian rationality: the probabilistic approach to human reasoning. Oxford University Press, Oxford

Ong AD, Weiss DJ (2000) The impact of anonymity on responses to sensitive questions. J Appl Soc Psychol 30(8):1691–1708. https://doi.org/10.1111/j.1559-1816.2000.tb02462.x

Pande BM (1965) Understanding culture. East Anthropol 18(1):34–43

Park E, Jin D, Del-Pobil A (2012) The law of attraction in human-robot interaction. Adv Robot Syst 9(35):1–7

Pettigrew TF (1979) The ultimate attribution error: extending Allport's cognitive analysis of prejudice. Pers Soc Psychol Bull 5(4):461–476

Pio-Abreu JL, Ferreira BR, Januario C (2015) The concept of meaning: the key to clarify the human cognition and psychopathology. Med Hypotheses 84(3):268–272

Polderman TJ, Stins JF, Posthuma D, Gosso MF, Verhulst FC, Boomsma DI (2006) The phenotypic and genotypic relation between working memory speed and capacity. Intelligence 34(6):549–560

Powers DM (1998) The total Turing test and the Loebner prize. In: Proceedings of the joint conferences on new methods in language processing and computational natural language learning, pp 279–280. https://dl.acm.org/citation.cfm?id=1603947

Premack D (2010) Why humans are unique. Perspect Psychol Sci 5(1):22–32

Preston B (1991) AI, anthropocentrism, and the evolution of 'intelligence.' Mind Mach 1(3):259–277. https://doi.org/10.1007/bf00351181

Provine RR (2001) Laughter: a scientific investigation. Penguin Books, New York

Psaltis C, Duveen G (2006) Social relations and cognitive development: The influence of conversation type and representations of gender. Eur J Soc Psychol 36(3):407–430

Putnam H (1994) Sense, nonsense, and the senses: an inquiry into the powers of the human mind. J Philos 91(9):445–517

Quiñonez Y, Ramirez M, Lizarraga C, Tostado I, Bekios J (2015) Autonomous robot navigation based on pattern recognition techniques and artificial neural networks. In: Ferrández Vicente J, Álvarez-Sánchez J, de la Paz López F, Toledo-Moreo F, Adeli H (eds) IWINAC 2015: bioinspired computation in artificial systems. Springer, New York, pp 320–329

Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP, Ng AY (2017). CheXNet: Radiologist-level pneumonia detection on chest x-rays with deep learning. arXiv. https://arxiv.org/abs/1711.05225

Reason J (1990) Human error. Cambridge University Press

Reeves B, Hancock J, Liu X (2020) Social robots are like real people: first impressions, attributes, and stereotyping of social robots. Technol Mind Behav 1(1)

Richards C, Bouman WP, Seal L et al (2016) Non-binary or genderqueer genders. Int Rev Psychiatry 28:95–102

Riding R, Cheema I (1991) Cognitive styles—an overview and integration. Educ Psychol 11(3–4):193–215

Ritchie G (2001) Current directions in computational humor. Artif Intell Rev 16(2):119–135

Roberts AL, Rosario M, Slopen N, Calzo JP, Austin SB (2013) Childhood gender nonconformity, bullying victimization, and depressive symptoms across adolescence and early adulthood: an 11-year longitudinal study. J Am Acad Child Adolesc Psychiatry 52(2):143–152

Robertson RE, Tran FW, Lewark LN, Epstein R (2017) Estimates of non-heterosexual prevalence: the roles of anonymity and privacy in survey methodology. Arch Sex Behav. https://doi.org/10.1007/s10508-017-1044-z

Rodrigo SM, Abraham JGF (2012) Development and implementation of a chat bot in a social network. Paper presented at Information Technology: New Generations (ITNG), 2012 Ninth International Conference, Las Vegas, NV, pp 751–755. https://doi.org/10.1109/ITNG.2012.147

Rogers C (2012) On becoming a person: a therapist's view of psychotherapy. Houghton Mifflin Harcourt, Boston

Rosen LD, Carrier ML, Cheever NA (2013) Facebook and texting made me do it: media-induced task switching while studying. Comput Hum Behav 29(3):948–958

Rutkin A (2015) Speech recognition AI identifies you by voice wherever you are. New Scientist, October 7. https://www.newscientist.com/article/mg22830423-100-speech-recognition-ai-identifies-you-by-voice-wherever-you-are/

Sadler P (1998) Psychometric models of student conceptions in science: reconciling qualitative studies and distractor-driven assessment instruments. J Res Sci Tech 35:265–296

Sainato M. (2015). Stephen Hawking, Elon Musk, and Bill Gates warn about artificial intelligence. Observer, August 19. https://observer.com/2015/08/stephen-hawking-elon-musk-and-bill-gates-warn-about-artificial-intelligence/

Sakai K, Sumioka H, Minato T, Nishio S, Ishiguro H (2013) Motion design of interactive small humanoid robot with visual illusion. Int J Innov Comput Inf Control 9(12):4725–4736. https://doi.org/10.1145/2350046.2350068

Schachter S, Singer J (1962) Cognitive, social, and physiological determinants of emotional state. Psych Rev 69(5): 379–399. https://doi.org/10.1037/h0046234

Schweizer P (1998) The truly total Turing test. Mind Mach 8:263–272. https://doi.org/10.1023/A:100822961954

Shipstead Z, Broadway JM (2013) Individual differences in working memory capacity and the Stroop effect: do high spans block the words? Learn Individ Differ 26:191–195

Shukla Shubhendu S, Vijay J, Vijay J (2013) Applicability of artificial intelligence in different fields of life. Int J Sci Eng Res 1(1):28–35

Smith A, Anderson J (2014) AI, robotics, and the future of jobs. Pew Research Center, 6

Smith DS, Juvonen J (2017) Do I fit in? Psychosocial ramifications of low gender typicality in early adolescence. J Adolesc 60:161–170

Smith R, Lane RD (2016) Unconscious emotion: a cognitive neuroscientific perspective. Neurosci Biobehav Rev 69:216–238

SolveMedia (2015). Solve Media. https://www.solvemedia.com

Stiefelhagen R, Fugen C, Gieselmann P, Holzapfel H, Nickel K, Waibel A (2004) Natural human-robot interaction using speech, head pose and gestures. Paper presented at the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan. https://doi.org/10.1109/IROS.2004.1389771
Stone-Gross B, Holz T, Stringhini G, Vigna G (2011) The underground economy of spam: a botmaster's perspective of coordinating large-scale spam campaigns. In: LEET'11 Proceedings of the 4th USENIX conference on large-scale exploits and emergent threats, p 4. https://www.usenix.org/legacy/events/leet11/tech/full_papers/Stone-Gross.pdf
SweetCAPTCHA (2015) SweetCAPTCHA: fun and human friendly CAPTCHA. SweetCaptcha. https://sweetcaptcha.com
Szollosy M (2017) Freud, Frankenstein and our fear of robots: projection in our cultural perception of technology. AI Soc 32:433–439. https://doi.org/10.1007/s00146-016-0654-7
Tannen D (1990) You just don't understand: Women and men in conversation. William Morrow, New York
Tay BT, Low SC, Ko KH, Park T (2016) Types of humor that robots can play. Comput Hum Behav 60:19–28
Taylor H (2016) Could you fall in love with this robot? CNBC, March 16. https://www.cnbc.com/2016/03/16/could-you-fall-in-love-with-this-robot.html
Taylor SE, Brown JD (1988) Illusion and well-being: a social psychological perspective on mental health. Psychol Bull 103(2):193
Ting-Toomey S, Chung LC (2005) Understanding intercultural communication. Oxford University Press
Tomasello M (2010) Origins of human communication. MIT Press, Cambridge
Tomasello M, Rakoczy H (2003) What makes human cognition unique? From individual to shared to collective intentionality. Mind Lang 18(2):121–147
Toomey RB, McGuire JK, Russell ST (2012) Heteronormativity, school climates, and perceived safety for gender nonconforming peers. J Adolesc 35(1):187–196
Towns MH (2014) Guide to developing high-quality, reliable, and valid multiple-choice assessments. J Chem Educ 91(9):1426–1431. https://doi.org/10.1021/ed500076x
Townsend JT, Ashby FG (1984) Measurement scales and statistics: the misconception misconceived. Psychol Bull 96(2):394–401
Treagust D (1988) Development and use of diagnostic tests to evaluate students' misconceptions in science. Int J Sci Educ 10:159–169
Trenholm S (2017) Thinking through communication: an introduction to the study of human communication. Routledge, New York
Tryon WW (2014) Clinical applications of principle 1: Unconscious processing. In: Tryon WW (ed) Cognitive neuroscience and psychotherapy. Academic Press, New York, pp 397–451
Turing AM (1950) Computing machinery and intelligence. Mind 59(236):433–460
Vaes J, Paladino MP (2010) The uniquely human content of stereotypes. Group Process Intergroup Relat 13(1):23–39
Vikram S, Fan Y, Gu G (2011). SEMAGE: a new image-based two-factor CAPTCHA. In Proceedings of the 27th annual computer security applications conference. ACM. Orlando, Florida, pp 237–246. https://doi.org/10.1145/2076732.2076766
Vincent J (2017) Send Scam emails to this chatbot and it'll waste their time for you. The Verge. https://www.theverge.com/2017/11/10/16632724/scam-chatbot-ai-email-rescam-netsafe
Vrij A, Blank H, Fisher R (2017) A re-analysis that supports our main results: a reply to Levine et al. Legal Criminological Psychol. https://doi.org/10.1111/lcrp.12121
Von Ahn L, Blum M, Hopper NJ, Langford J (2003). CAPTCHA: using hard AI problems for security. In: International conference on the theory and applications of cryptographic techniques. Springer, pp 294–311. https://doi.org/10.1007/3-540-39200-9_18
Wei TE, Jeng AB, Lee HM (2012) GeoCAPTCHA—a novel personalized CAPTCHA using geographic concept to defend against 3rd party human attack. In: Performance computing and communications conference (IPCCC), 2012 IEEE 31st international. IEEE, pp 392–399
Weizenbaum J (1966) ELIZA—a computer program for the study of natural language communication between man and machine. Commun ACM 9(1):36–45
West DM (2015) What happens if robots take the jobs? The impact of emerging technologies on employment and public policy. Inside Politics. https://www.insidepolitics.org/brookingsreports/robots.pdf
Wickens CD, Hollands JG, Banbury S, Parasuraman R (2015) Engineering psychology and human performance. Routledge
Wierzbicka A (1986) Human emotions: universal or culture-specific? Am Anthropol 88(3):584–594
Withers LA, Vernon LL (2006) To err is human: embarrassment, attachment, and communication apprehension. Personality Individ Differ 40(1):99–110
Wood JT (1997) Gendered lives: Communication, gender and culture, 2nd edn. Wadsworth, Belmont
Woolley S, Gorbis M (2017) Social media bots threaten democracy. But we are not helpless. The Guardian, October, 16. https://www.theguardian.com/commentisfree/2017/oct/16/bots-social-media-threaten-democracy-technology
Yan D, Rupp AA, Foltz PW (2020) Handbook of automated scoring theory into practice. CRC Press, Taylor & Francis Group
Young RM (2001) Chapter 1: locating and relocating psychoanalytic ideas of sexuality. In: Sexuality: psychoanalytic perspectives. Brunner-Routledge, pp 18–34
Yu Z (2017) Situated intelligent interactive systems. Doctoral dissertation. Carnegie Mellon University, Language Technologies Institute. https://www.lti.cs.cmu.edu/sites/default/files/yu%2C%20zhou%20-%20CMU-LTI-01-018.pdf
Zabihimayvan M, Sadeghi R, Rude HN, Doran D (2017) A soft computing approach for benign and malicious web robot detection. Expert Syst Appl 87:129–140
Zarkadakis G (2016). In our own image: Savior or destroyer? The history and future of artificial intelligence. Pegasus.
Zerega B (2017) The Turing Test is tired. It's time for AI to move on. Huffington Post, October 18. https://www.huffingtonpost.com/entry/the-turing-test-is-tired-its-time-for-ai-to-move_us_59bfae52e4b02c642e4a187a
Zhao Y, Montoro R, Igartua K, Thombs BD (2010) Suicidal ideation and attempt among adolescents reporting "unsure" sexual identity or heterosexual identity plus same-sex attraction or behavior: forgotten groups? J Am Acad Child Adolesc Psychiatry 49:104–113
Zlotowski J, Sumioka H, Nishio S, Glas DF, Bartneck C, Ishiguro H (2016) Appearance of a robot affects the impact of its behavior on perceived trustworthiness and empathy. J Behav Robot 7(1):55–66. https://doi.org/10.1515/pjbr-2016-0005