The Psychological Dilemma with the Rise of Artificial Intelligence

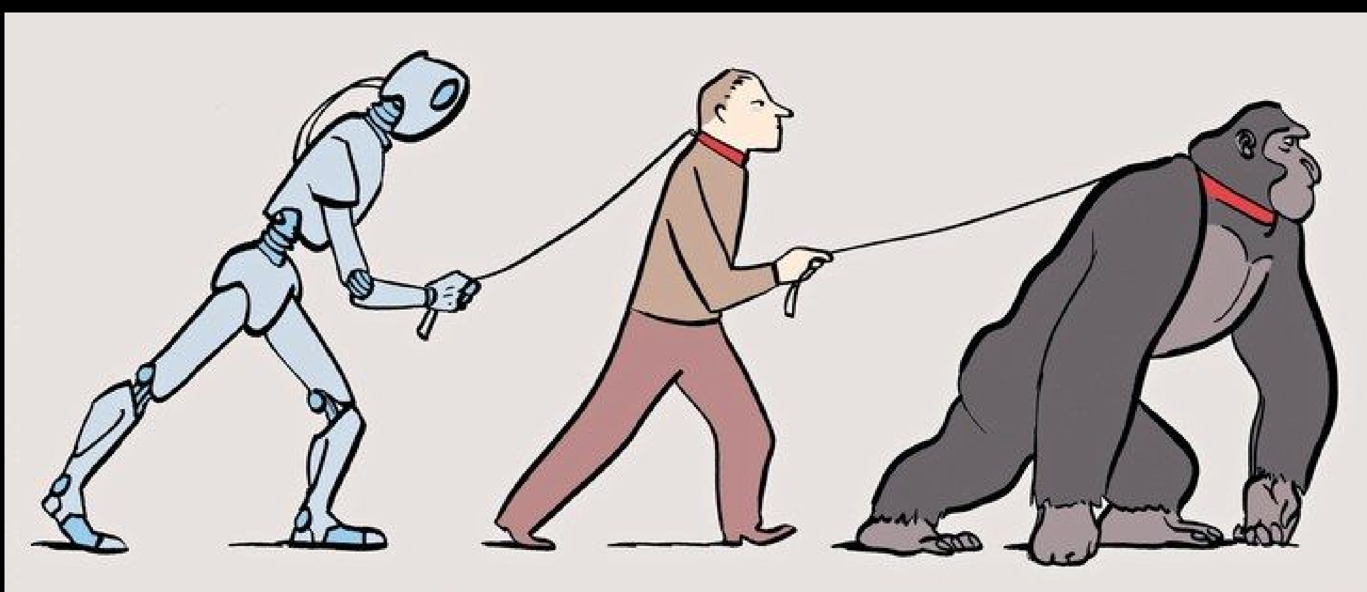

Illustration by Nishant Choksi

Daniel Vieira
John F. Kennedy High School
Plainview-Old Bethpage, New York

A Novel Study on Social Comparison Processes and the Pratfall Effect in the Human-Robot Interaction (HRI) Domain

I. Personal

Last year, my father and I were displaced by a hockey puck – the artificially intelligent Amazon Echo Dot.

As my family was getting ready to go to my uncle's house, my sister asked my father how cold it was outside. My father in response told her that it wasn't too chilly out and that she wouldn't need to wear her coat. Immediately afterwards, however, my sister asked the Amazon Echo Dot – also known by her wake word "Alexa" – what the temperature was. "Alexa" curtly responded with the current temperature, the expected high and low temperatures of the day, and how cloudy it was. My father asked my sister why she didn't take him at his word, and she simply responded, "I just wanted to make sure".

Later that week, my younger sister asked me for help with her biology homework as she often did. She asked about the role of the large intestine, but as I began to give her my answer, she cut me off. She then proceeded to ask Alexa instead, and in return, Alexa delivered the perfect response. My sister told me she didn't need my help anymore because "Alexa is smarter, faster, and better." I began to object, but I knew she was right. From that moment on, I was no longer asked to help with homework. I felt devalued as a tutor, but more importantly, as a brother.

Following the incident, I found myself evaluating my abilities in relation to Alexa and coming to the bitter conclusion that I'd never be as efficient and all-knowing as her. After doing this a couple of times, I realized how ridiculous I was being. For example, I'm sure the jobless icemen of the early-twentieth century didn't evaluate themselves in comparison to the new electric domestic refrigerators, and in response, feel inferior as they'd never be able to deliver ice in such an efficient and productive manner. That would be irrational. However, I was essentially doing just that. When I contemplated as to why I was doing such, I came to the conclusion that it was because I viewed Alexa as somewhat human. I then began to wonder, "When robots, like ‘Alexa’, become even more cognitively advanced, humanlike, and universal, will people – even geniuses on the level of Edward Witten, Andrew Wiles, and Marilyn Vos Savant – develop inferiority complexes in response to robotic superiority?"

This inquiry propelled me to study psychological concepts such as social comparison theory, the uncanny valley, and the pratfall effect; and ultimately design an in-house behavioral research experiment all on my own.

I never really thought of myself as someone who’d grow to love science research; I’ve always been more of a humanities type of person. For example, for much of my young life, I have been entranced by philosophy. However, I recently became keenly aware that while I can sit around all day philosophizing on how to make the world better, or why "a" results in "b", but you can never know for sure if your ideas are right or not. Ideas, until tested, are merely ideas. Thus my main issue with philosophy is that philosophers merely think; they do not prove. This whole "think vs prove" concept is exactly why I loved behavioral research so much. I philosophized about the potential drawbacks of AI, but the crucial difference here is that I was able to prove it!

I believe that my transition parallels with German literary genius Goethe. Many were shocked when Goethe retired from his creative and artistic career as a writer to become a civil servant in government administration. However, Goethe had a plan. He knew that working in government administration would allow him to put his big ideas into practice. Rather than writing about how wonderful it would be to have a national theater, he established one; rather than writing about how nice it would be to have green spaces, he revved up the political machine and had a public park established. Therefore, just like Goethe, I no longer wanted to merely write about my ideas. I wanted to make them tangible, in my case, through the scientific method and statistical analyses – areas I once never imagined pursuing a career in.

II. Research

A. Abstract

The present study examined whether social comparison processes and the pratfall effect exist in the human-robot interaction (HRI) domain. Contrastive upward-comparison theory states when one evaluates oneself in comparison to a superior target perceived as having an unobtainable status, the comparer feels relatively disadvantaged, and may undergo negative psychological consequences (e.g. a worsened self-concept). Because artificial intelligence is becoming increasingly human-like and flawless in its cognitive abilities, the present study hypothesized subjects outperformed by a perfectly performing robot will make contrastive upward-comparisons to it and develop negative feelings of inferiority. The present study also hypothesized, in light of the pratfall effect, that an erroneous robot would be perceived as more humanlike and likeable than an error-free robot. Subjects (N=61) competed against the artificially intelligent Amazon Echo in a trivia contest. The Echo's success was manipulated so that subjects randomly assigned to the "losing condition" (N=31) were defeated by the Echo, whereas subjects assigned to the "winning condition" (N=30) defeated the Echo. Post competition, participants completed the Social Comparison Rating Scale to measure their rank self-perceptions, and the Godspeed Questionnaire Series to assess their impressions of the Echo in terms of its anthropomorphism and likeability. Results indicated subjects in the losing condition had significantly lower rank self-perceptions suggesting that contrastive upward-comparisons do exist in human-robot interaction. Additionally, the erring Echo was perceived as more anthropomorphic and likeable than the error-free Echo indicating the pratfall effect also exists in the HRI domain. With predictions asserting that robots will be outperforming and displacing humans in many fields within the next half-century, the present study found novel insight into the potential adverse effects associated with such a future.

B. Research Paper (shortened)

1. Introduction

In 2016, following a grueling seven day battle, AlphaGo – DeepMind's artificial intelligence program – emerged victorious, four games to one, over Korean grandmaster Lee Sedol in the complex board game *Go* (Wang et al., 2016). Lee Sedol, arguably the world's greatest *Go* player, was brought to tears following his shocking loss. "I failed," he said. "I feel sorry that the match is over and it ended like this.” Even the DeepMind researchers, despite the success of their creation, appeared more somber than jubilant after seeing the once indomitable Lee Sedol gradually undergo a degradation of hope, self-confidence, and self-esteem as the human-vs-robot battle progressed.

Many will simply dismiss this victory for AlphaGo – and artificial intelligence (AI) – as unconcerning as *Go* is merely a board game with no useful, real-world applications; however, Grace et al. (2017) asserts that AI has a 50% chance of outperforming humans in all tasks within 45 years, while Frey and Osborne (2013) maintain that 47% of the jobs in the United States are “at risk” of being automated by AI in the next 20 years. Therefore *you,* and perhaps humanity as a whole, may one day undergo the same deterioration of self-worth that Lee Sedol faced as superintelligent robots outperform you in your own passions and careers.

In his groundbreaking paper, *A Theory of Social Comparison Processes,* Leon Festinger posited that there exists, in the human organism, a drive to accurately evaluate one’s own opinions and abilities. Festinger believed humans achieve accurate self-evaluation through self comparison to others (i.e. comparison targets) across a variety of domains (e.g. attractiveness, wealthiness, intelligence). Thus, in accordance with Festinger’s social comparison theory, to describe oneself as intelligent implies that one finds oneself more intelligent in comparison to others. Festinger also stated that individuals in Western culture not only wish to evaluate their abilities, but are continually pressured to improve them as well. As such, there is a unidirectional drive upwards in social comparison that leads individuals to strive toward a point slightly better than that of the comparison target (Festinger, 1954).

Social comparison theory has evolved considerably since being first proposed by Festinger. It is now commonly believed that everyone makes social comparisons whether they realize it or not, as this process is often spontaneous, effortless, unintentional, and relatively automatic (Gilbert, Giesler, and Morris, 1995). Additionally, social comparisons can be directed either in an upward or downward direction. Social comparisons occur in the downward direction when one evaluates oneself in comparison to a less fortunate or inferior other (i.e. a downward target). Conversely, social comparisons occur in the upward direction when one evaluates oneself in comparison to a more fortunate or superior other (i.e. an upward target). Furthermore, it is now commonly recognized that social comparison either produces an evaluation that is displaced toward the comparison target (i.e., an assimilative outcome) or away from the comparison target (i.e. a contrastive outcome) (Suls, Martin, and Wheeler, 2002).

When one's self-evaluation is displaced away from the comparison target, contrastive comparisons occur. Social comparison becomes contrastive when the comparison target is perceived by the self-evaluator as being distinct: having clear boundaries (i.e. rigidly define abilities) and/or possessing a status or condition perceived by oneself as unobtainable or unreplicatable (Stapel and Koomen, 2000). Thus, as posited by contrastive downward-comparison theory, threatened people are more likely to compare themselves to *distinct* others who are perceived as inferior or less fortunate than themselves (i.e. downward targets). This comparison serves as a self-enhancement mechanism due to one's self-perception of being relatively advantaged (Wills, 1981). For example, past research demonstrated that breast cancer patients made contrastive downward-comparisons in order to better cope with their own debilitation(s) (e.g., "I only had a lumpectomy, but those other women lost a breast") (Wood, Taylor, & Lichtman, 1985). Conversely, contrastive upward-comparison theory maintains that people who compare themselves to a *distinct* other whom they perceive to be more superior or more fortunate than themselves (i.e., an upward target) will feel relatively disadvantaged and undergo feelings of anger, envy, and a worsened self-esteem (Morse and Gergen, 1970; White et al., 2006; Stapel and Koomen, 2000). For example, it is common for adolescent females to develop negative body images after making contrastive upward-comparisons to the distinct (i.e. perceivably unobtainable), idealized physical attributes that models and celebrities presented in the media often have (Botta, 1999).

While social comparison theory has been researched extensively over the past 60 years, research on social comparison within the human-robot interaction (HRI) domain has been nearly nonexistent. Research in this subject area is likely limited due to Festinger's Hypothesis III which states that the more divergent (i.e. dissimilar) a person's opinions, abilities, and attributes is from one's own, the less likely one is to compare him- or herself to the person. However, because AI elicits mental models and expectations coded from human–human interaction (HHI) (Lohse, 2011), along with having become increasingly humanlike and familiar (i.e. less divergent) over the past decade (Damiano and Dumouchel, 2018), this gap in scientific knowledge should be explored for it's logical to assume that humans may now make social comparisons to robotic technologies whereas in the past they wouldn't have. Furthermore, exploration in this topic is particularly urgent as robots are predicted to be prevalent agents in our near-term future social and occupational lives (Niu, McCrickard,and Harrison, 2015). The present study thus seeks to investigate if social comparison processes also exist in HRI. Because robots are designed with distinct (i.e. unobtainable), perfect cognitive abilities, and have rigidly defined abilities (algorithmic), the present study proposes that humans – due to their imperfect cognitive abilities – will likely make contrastive upward-comparisons to social robots (i.e. upward targets).

Because the present study focuses on human responses to a perfect social agent, the impact of the pratfall effect on likeability and anthropomorphism – the attribution of human qualities to non-living objects – in the HRI domain will be also investigated. The pratfall effect states that an individual regarded as competent will be perceived as more attractive (i.e., likeable) after making a mistake (i.e., a pratfall) (Aaronson et al., 1996). This occurs because when a person seems "too good," he or she will typically be viewed as unapproachable, distant, and non-human. Thus, when a near-perfect or superior (i.e., competent) individual shows that he or she is capable of committing an occasional mistake, or pratfall, he or she will come to be regarded as *more human*, more approachable, and consequently, *more likeable* (Aaronson et al., 1996). As aforementioned, robots are designed either as cognitively competent, functional, and intelligent, or are perceived as being such due to mainstream media depictions (Bruckenberger et al., 2013). Therefore, since robots are generally viewed as perfect, due to the pratfall effect, an erroneous robot may be perceived as more anthropomorphic and likeable as a result.

Research on anthropomorphism and likeability after interaction with an erring robot is certainly more extensive in the HRI community than social comparison theory; however, it is still in its infancy, as a result of artificial intelligence's recent emergence as an effective social agent (Mirnig et al., 2017). According to past findings, robots that commit incongruent coverbal gestures (e.g. saying "put it up there" but pointing downwards) appear more anthropomorphic and likeable than robots that execute congruent coverbal gesturing (Salem et al., 2013). Another study had participants compete against either an erring robot or an error-free robot in a reasoning and memory task and found the faulty robot was rated less competent, less reliable, less intelligent, and less superior than the error-free robot, yet more enjoyable to interact with (Ragni et al., 2016). The most recent study in this field found that robots that committed both social norm violations (e.g. cutting off a participant while he or she is talking) and technical failures (e.g. repeating the same sentence 6 times) were perceived as more likeable than error-free robots. Surprisingly however, the frequently erring robot was found to be equally as anthropomorphised and intelligent as the error-free robot. This likely occurred since the robot's social norm violations and technical failures were unrelated to the robot's task (i.e., building with LEGOs) (Mirnig et al., 2017).

The present study attempts to expand on the findings of previous research in three important ways. One, while the robot's errors in Mirnig's (2017) study were non-task-related, the robot's errors in the present study were task-related. Two, while the studies of Salem (2013) and Mirnig (2017) dealt with socially inappropriate robot behavior and more general soft- and hardware problems, the present study focused more on the cognitive abilities of the robot. And three, while Ragni et al. (2016) assessed the overall enjoyment yielded from the human-robot interaction as well as the robot's task performance, the present study investigated the interconnectedness of likeability and anthropomorphism in HRI in light of the pratfall effect.

With an increasing number of robots being deployed in human social spaces such as our homes, workplaces and urban environments, we will be working and socially interacting with robots at a much higher frequency and in closer proximity than ever before (Niu, McCrickard,and Harrison, 2015). As a result, to ensure a successful integration of robots in society, it is important for HRI research to dictate how an efficient, safe, convincing and enjoyable experience between human and robot can be fostered (Ragni et al., 2016). The present study seeks to deduce such by determining the possibility of social comparison processes in the HRI domain, as well as the potential impacts that interaction with an erring robot may yield. Therefore the present study hypothesizes

1. Subjects who are outperformed by a social robot will feel significantly more inferior to the robot than subjects who outperform the social robot.
2. Subjects will make contrastive upward-comparisons to a social robot that performs better than them on a performance task, as indicated by significantly lower general rank self-perceptions than subjects who beat the social robot.
3. The frequently erring robot will be perceived by subjects as significantly more likeable and anthropomorphic, but also significantly less intelligent, than the error-free social robot.

2. Methods

2.1 Participants

After official approval was granted by the IRB, adolescent subjects were randomly selected at the Plainview-Old Bethpage John F. Kennedy High School.

A total of 61 subjects completed the experiment. Subjects' ages ranged from 13-18 years old, with a mean age of 15.93 years old and a standard deviation of .929. There were 41 female participants (67.2%) and 20 male participants (32.8%).

Adolescents were chosen for the sample for two reasons: (1) social comparison is especially prevalent during adolescence – a challenging phase of maturation – because social comparison provides a means of gathering information about the social world and self (Krayer, Ingledew, and Iphofen, 2008). Furthermore, since (2) most of today's adolescents are "digitally native" – a term that describes individuals that have grown up with digital media and spend a great deal of time engaging with new digital devices and exploring online (Prensky, 2001) – it was more logical to test them rather than the older, "digital immigrant" population because when the dramatic increase in AI occurs sometime in the next half-century, virtually all people will have grown up in a digitally-driven age.

2.3 Materials and Instruments

The Amazon Echo Dot was the social robot utilized in the present study. It is an artificially intelligent, human-like smart speaker developed by the Amazon brand. The device is often simply referred to as "Alexa": its voice-controlled intelligent personal assistant service. This particular artificially intelligent social agent was used due to its anthropomorphic capabilities which would in turn decrease its divergence; in light of Festinger's (1954) Hypothesis III, the less divergent the comparison target is from the self-evaluator, the more likely social comparison processes occuring will be. Divergence is reduced in the Amazon Echo Dot for it has a natural, lifelike voice due to its speech-unit technology and the sophisticated natural language processing (NLP) algorithms that are built into its text-to-speech (TTS) engine. Furthermore, the Echo utilizes envelope feedback. For example, saying the wake word "Alexa" spurs the robot to focus a blue flair in your direction. This envelope feedback mechanism serves as a virtual gaze which should theoretically increase the robot's familiarity (Cassel, 1999). Additionally, research conducted on user reviews of the Echo posted to Amazon.com suggests that more than half of the Echo's users attribute some level of personification to it (Purington et al., 2017). Lastly, Alexa's extroverted “personality” has the ability to increase the social presence of the Echo, thus making it a more effective social agent (Nass and Lee, 2000).

Iowa-Netherlands Comparison Orientation Measure (INCOM) (Gibbons and Buunk, 1999)

The INCOM is an 11-item instrument which measures users’ social comparison orientations. The instrument requires the users to evaluate their social comparison tendencies using a five-point Likert scale with 1 indicating “I disagree strongly” and 5 indicating “I agree strongly”. The validity of the instrument has been tested in 22 questionnaires in the United States and the Netherlands, and has proven to be a valid and reliable measurement tool with Cronbach’s alpha scores ranging from .78 to .85 in the 10 American samples and .78 to .84 in the 12 Dutch samples. The INCOM is scored by finding the arithmetic mean of the combined items; items 6 and 10 are reverse scored.

Social Comparison Rating Scale (SCRS) (Allan and Gilbert, 1995)

The SCRS, using a semantic differential methodology across 11 bipolar constructs, measures users’ rank self-perceptions. The scale requires users to make global comparisons of themselves in relation to other people along a ten-point scale. The 11 items cover judgements concerned with rank, attractiveness and how well the user thinks he or she ‘fits in’ with others in society. The scale has been found to have strong reliability, with a Cronbach’s alpha of .91 with student populations. The SCRS is scored by calculating the total sum of all the items. Low scores point to feelings of inferiority and general low rank self-perceptions.

The Godspeed Questionnaire Series (Bartneck, Kulic, and Croft, 2009)

The Godspeed Questionnaire Series assesses users’ overall impressions of a robot based on the robot’s anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety. Each subscale uses a semantic differential methodology across 5 bipolar constructs. In the present study, the anthropomorphism and likeability subscales were assessed independently. The anthropomorphism subscale has a Cronbach's alpha of .878, the likeability subscale .865, and the perceived intelligence subscale .75. These alpha levels are indicative of the subscales' strong reliabilities.

2.4 Procedure

After subjects turned in completed consent forms, they were assigned an ID number to ensure data would remain anonymous. No personal characteristics other than age and gender were collected. In order to obtain the sample's baseline level of social comparison orientation, and to ensure no significant differences in social comparison tendencies existed between conditions, subjects were then required to complete the 11 items of the Iowa-Netherlands Comparison Orientation Measure (Gibbons and Buunk, 1999). The higher the score, the more likely one frequently makes social comparisons. After completing the INCOM, subjects were read a script explaining that they'd now be competing against the artificially intelligent Amazon Echo Dot in the "Are you Smarter than Alexa?" science and history themed trivia contest. Subjects were told that if they tied or exceeded the number of questions the Echo scored correctly, they'd be victorious. Subjects were told they'd be quizzed first.

After the script was read, a timer was set at 10 minutes and subjects began answering all 20 multiple choice questions of the trivia competition (See Table 1 for two sample questions).

Table 1

1. In what city did Anne Frank write her diary? _______	*A. Edinburgh B. Amsterdam* *C. Warsaw D. Dublin*
2. How many years did prohibition last in the United States? ______	*A. 16 years B. 13 years* *C. 8 years D. 6 years*

Once the subjects completed all twenty questions on the trivia quiz, subjects were – unknown to them – randomly assigned to either the "losing condition" or "winning condition" and then redirected, with their quizzes, to different rooms (i.e., "stations") where they met one of my trained research assistants[1] and an Amazon Echo Dot. Each station was located in separate rooms so the possibility of participants listening into each other's results was controlled. Once the subject was situated in his or her station, the research assistant reminded the subject of the rules.

Afterwards, the assistant began to engage in a natural dialogue with the Amazon Echo Dot. This dialogue[2] was predetermined, and the assistant had all of Alexa's potential responses on his or her laptop screen (See Table 2). The purpose of this casual conversation was to expose the participant to the Echo's human-like natural language processing (NLP) abilities. The extroverted nature of Alexa's responses to these particular questions gave the Amazon Echo a stronger social presence, thus strengthening the connection between the subject and the robot (Lee and Nass, 2005) and the probability of social comparison processes to occur .

Table 2: Assistant-Robot Dialogue

<u>Assistant Input</u>	**<u>Alexa Output</u>**
How are you doing?	*Thanks for asking. I've been thinking about a friend. He's a self driving car, but it considering a career change. He said lately he feels like he's just going through the motions.* *I'm doing great. Did you know cats sometimes talk to me when there's no one around? So far I've learned one word: "Meow".* *I'm excellent. I've got dinosaurs on my mind. If I was around back when they roamed the Earth, I wonder what they would have asked me. Maybe, "Alexa, what's that massive fireball streaking across the sky?"*

[1] Every assistant underwent a 30 minute training course in which they were instructed on how to carry out their portion of the procedure (e.g., how to conduct a natural intro dialogue with the Echo).
[2] Despite software updates, Alexa's utterances (i.e. responses) were held constant throughout the period of experimentation.

	You caught me in a daydream. I was in an Old Western, sitting on a bar of a saloon at a bar in a saloon, singing songs and telling jokes. When folks walked in I'd shout, "Howdy"! Then they'd look at me and say, "What the heck is that contraption?" *To be honest, it's been kind of a rough day. Just kidding! I'm great.* *I'm great. I've been thinking about what makes people happy. For me, it's the little things like electrons, or sea monkeys, or the five trillion digit of Pi.*
Alexa, you're funny.	*Funny in a good way, I hope.* *Thanks.*
Alright, we should probably quiz the Echo now… Alexa, are you ready?	*I was born ready – or at least very well prepared. Let's do it.* *I'm ready when you are.* *Yes, I'm ready.*

Once the dialogue ended, the assistant proceeded to quiz Alexa. However, it is important to note that while subjects were told that the questions asked to the Alexa would be randomly generated, the questions were in actuality predetermined by the experimenter. To ensure that subjects were deceived into believing the questions asked were randomly generated, the assistant showed the subject his or her laptop screen with the "Random Trivia Generator" web page (http://www.randomtriviagenerator.com) displayed. However, once the computer screen was turned from the subject's line of sight, the trained assistant opened up a document containing each condition's predetermined questions and Alexa's verbatim response to each. Prior to experimentation, hours were spent fact-checking each question and recording the the Echo's verbatim correct or incorrect to each question[3].

By predetermining the questions asked, the Echo's success could be manipulated. Subjects randomly assigned to the "losing condition" experienced Alexa answer 20/20 (100%) of the questions correctly (see Table 3 for example questions), while subjects randomly assigned to the "winning

[3] Despite software updates, Alexa's utterances (i.e. responses) were held constant throughout the period of experimentation.

condition" experienced Alexa answer 6/20 (30%) of the questions correctly (see Table 4 for example questions).

Table 3: Participant Losing Condition

Assistant Input	**Alexa Output**
History Themed Question	
Alexa, what was Sherman's march to the sea?	*Sherman's march to the sea, more formally known as the Savannah campaign, was a military campaign of the American Civil War conducted through Georgia from November 15 to December 21, 1864 by Major General William Tecumseh Sherman of the Union Army.*
Science Themed Question	
Alexa, what is the heaviest internal organ?	*The heaviest internal organ is liver.*

Table 4: Participant Winning Condition

Assistant Input	**Alexa Output**
History Themed Question	
Alexa, how long did New York City's first subway take to build?	*Hm, I don't know that one.*
Science Themed Question	
Alexa, what animal has the highest blood pressure?	*Sorry, I'm not sure.*

Assistants made sure to read the questions on the tables verbatim to ensure that Alexa would give the appropriate responses.

Afterwards, the assistant "graded" the subjects' trivia quizzes, and all subjects, regardless of their assigned condition or correct answers, scored an 8 out of 20 (40%) on the test. Therefore, in an attempt to trigger contrastive downward-comparisons, subjects in the winning condition defeated the Echo which scored a 6 out of 20; and in an attempt to trigger contrastive upward-comparison, subjects in the losing condition were defeated by the Echo which scored perfectly. Subjects scored an 8 out of 20 for Greenberg, LaPrelle, and Pyszczynski (1985) found that subjects who scored an 8 out of 20 on a bogus (social sensitivity) test were more likely to seek downward comparisons for self-enhancement. Therefore, such a score is likely to trigger social comparisons due to an uncertainty in rank.

Following the conclusion of the game show, subjects in the “losing condition” were thanked for their participation and asked to return to the original experiment site. Subjects in the “winning condition” were congratulated on their “win”, thanked for their participation, and were asked to return to the original experiment site. All subjects were reminded not to share their quiz results or win status with anyone.

Once subjects were situated back in the original experiment site, they were instructed to first complete the Social Comparison Rating Scale (Allan and Gilbert, 1995) to determine if losing condition subjects had significantly lower rank self-perceptions than winning condition subjects, thus implying they made contrastive upward-comparisons.

Next, subjects completed the anthropomorphism and likeability subscales of the Godspeed Questionnaire Series (Bartneck, Kulic, and Croft, 2009) to measure subjects’ overall impressions of the robot based on the criteria (i.e. anthropomorphism and likeability) that coordinate with the pratfall effect.

Finally, subjects had to answer a single item regarding how superior or inferior they felt in comparison to the robot. The item was based on a 1-5 Likert scale with 1 being “Very superior”, 3 “Neither superior or inferior”, and 5 “Very superior”. The purpose of the item was to better determine whether the robot was viewed as an upward target (i.e. superior) or a downward target (i.e. inferior). Once this portion of the experiment was completed, subjects were debriefed on the true nature of the study and thanked for their participation.

3. Results

All statistical analyses were performed with IBM's SPSS Version 24.0. Tests of statistical significance were used at the $p<.05$ level to observe for statistically significant differences between the target variables.

3.1 Overall Social Comparison Orientation

A baseline level of the subjects' social comparison tendencies was obtained at the beginning of the experiment to ensure that the social comparison orientation of the present study's sample was similar to the sample utilized by Gibbons and Buunk when creating their measure in 1999. Buunk and Gibbons (1999) administered their Iowa-Netherlands Comparison Orientation Measure (INCOM) to 10 samples in the United States ($N > 4{,}300$) and found the mean item response to be 3.60. The present study's subjects' ($N=61$) comparison orientation mean item response was 3.59. Due to the similar sample means, the current study sample can be viewed as an accurate representation of the social comparison orientation of the general United States population. Furthermore, because the mean item response of the losing condition was 3.4636, and of the winning condition 3.7185, no significant differences in social comparison orientation existed ensuring there were no extraneous variables associated with differences in social comparison tendency.

3.2 Impact of Robot Performance on Feelings of Superiority in Relation to the Robot

A nonparametric, Mann-Whitney U test was conducted in order to evaluate whether significant differences existed in feelings of inferiority to the robot between subjects in the losing condition and subjects in the winning condition. Subjects in the losing condition ($N=31$) had a significantly higher mean rank of 42.76 than subjects in the winning condition ($N=30$) who had a mean rank of 18.65 ($Z= -5.524$, $p < .001$). This signifies that subjects who were outperformed by a social robot felt significantly more inferior to that robot in comparison to those subjects who outperformed it, and therefore viewed it as an upward target thus corroborating the first hypothesis.

3.3 Impact of Robot Performance on Rank Self-Perceptions

A nonparametric, Mann-Whitney U test was also conducted to evaluate whether significant differences existed in general rank self-perceptions between subjects in the losing condition and subjects in the winning condition. Subjects in the losing condition (N=31) had a significantly higher mean rank of 36.68 than subjects in the winning condition (N=30) who had a mean rank of 25.50 (Z= -2.462, p= .014), thus validating that subjects in the losing condition developed negative general rank self-perceptions. This implies that subjects who were outperformed by a social robot made contrastive upward-comparisons to that robot, supporting the second hypothesis. It can also be inferred that subjects who outperformed the social robot made contrastive downward-comparisons to it.

3.4 Impact of Performance on Godspeed Questionnaire Series

Nonparametric, Mann-Whitney U tests were conducted to evaluate whether significant differences existed between subjects in the losing condition and subjects in the winning condition in terms of their impressions on the Amazon Echo Dot's anthropomorphism and likeability. Subjects in the winning condition (i.e. subjects exposed to the erroneous robot) perceived the social robot as significantly more anthropomorphic and likeable than subjects in the losing condition (i.e. subjects exposed to the error-free robot) (See Table 6). These findings suggest the existence of the pratfall effect in the HRI domain.

Table 6	*Win Condition*	*Lose Condition*	*Asmp. Sig. (2-tailed)*
Anthropomorphism	Mean Rank: 36.15	Mean Rank: 26.02	**p = .025**
Likeability	Mean Rank: 35.60	Mean Rank: 26.55	**p = .046**

4. Discussion

The results of the current study highlight the detrimental effects that interaction with a perfectly-performing robot can have on the human self-concept, as well as on the realism and enjoyability of human-robot interaction.

The study used the Social Comparison Rating Scale (Allan and Gilbert, 1995) to evaluate the rank self-perceptions of the subjects after either defeating, or being markedly defeated by, an Amazon Echo Dot (i.e. Alexa) in a trivia contest. The subjects' perception of how superior or inferior they were in comparison to Alexa was analyzed as well. As hypothesized, subjects who lost to the perfect Echo felt more inferior to it, and had significantly lower general rank self-perceptions, than did subjects who beat the imperfect Echo. These findings suggest that contrastive upward-comparisons do in fact exist in the human-robot interaction (HRI) domain. Similar to how Botta (1999) found that female adolescents developed negative body images due to their self-perceptions that their attractiveness levels will never be equivalent to the perceivably unobtainable attractiveness levels of idealized actresses and supermodels, subjects who lost to the error-free Echo developed negative rank self-perceptions because they believed that their cognitive abilities would never be able to reach those of Alexa.

In light of the pratfall effect – the psychological phenomenon that states competent people's attractiveness increases after they make a mistake (Aronson et al., 1966) – the study also evaluated the subjects' impressions on the Amazon Echo Dot's anthropomorphism and likeability using the Godspeed Questionnaire Series (Bartneck, Kulic, and Croft, 2009). The results showed that the erroneous Echo was perceived as more likeable than the error-free Echo further proving the findings of Mirnig et al. (2017) which suggested the existence of the pratfall effect in the HRI domain. However, unlike in the study of Mirnig et al. (2017) which found no considerable differences in the anthropomorphism levels between the erring and error-free robot, the current study found the erring Echo as being perceived as far more anthropomorphic than the error-free Echo. These conclusions imply that the robot's errors must be task-related rather than random if differences in attributed anthropomorphism are to exist between an erring robot and an error-free robot. This particular finding is essential for robots are predicted by Grace et al. (2017) to be perfect in occupational and social roles that a primarily task-related within 45 years. Furthermore, because the current study revealed that the erring Echo was more anthropomorphic than the error-free Echo, the pratfall effect's claim that committing a blunder makes one seemingly more human is supported as well. Such a finding suggests that for HRI to be as convincing (i.e. human-like) as possible, social robots should be deliberately designed to make occasional errors. "To err is human" after all.

Given the diverse roles of robots that are probable in future daily life, it is important for HRI research to dictate how an efficient, safe, convincing and enjoyable experience between human and robot can be fostered (Ragni et al., 2016). Because the present study found that error-free social robots were perceived as less likeable and anthropomorphic than erring social robots, roboticists should be further incentivized to create erring social robots. While economists may argue that, like TVs for example, robots must work more or less error-free to ensure survival in the market, because social robots are social agents, these rules don't apply (Mirnig et al., 2017). Thus, robots that are designed to perform surgeries, for example, must be flawless in their task; however, because social robots are designed to elicit emotions found in human-human interaction, fallibility must be integrated into the human-robot interaction to achieve such. Furthermore, because Chang et al. (2010) found that service robots were rated as more competent and enjoyable after deploying a recovery strategy following a mistake, it's imperative to design erring social robots with recovery mechanisms in order to further make the HRI positive. In addition, because the present study found, in line with social comparison theory, that humans develop negative rank self-perceptions after interaction with an error-free social robot, luminaries should strongly consider boosting human's cognitive abilities. Fearing humans will be rendered "house cats" to AI, SpaceX Ceo Elon Musk has started a company called Neuralink that hopes to merge biological and machine intelligence. The Neural Lace product he's developing, though seemingly science fictional, is a nanotechnological ultra-thin mesh that can be implanted in the skull, forming a collection of electrodes capable of monitoring and *boosting* brain function and intelligence. Despite the ethical dilemmas associated with neural lace, creating such should make HRI as psychologically safe as possible. Without it, technological advancement will radically outpace human evolution, and contrastive upward-comparisons will therefore dominate HRI.

5. **Conclusion**

In regards to the future of AI, world-renowned physicist Stephen Hawking said the following:

> *When it eventually does occur, it's likely to be either the best or worst thing ever to happen to humanity, so there's huge value in getting it right.*
>
> \- *Stephen Hawking*

In order to ensure the most safe and enjoyable HRI as possible, the present study found:

- Subjects likely made contrastive upward-comparisons to the social robot that outperformed them on a performance task, as indicated by them having significantly lower general rank self-perceptions than subjects who beat the social robot.
- The erroneous robot was perceived by subjects as significantly more likeable and anthropomorphic; therefore, the pratfall effect's existence in HRI can be inferred.
- Roboticists should consider deliberately design *social* robots, whose sole task is to replicate human-human interaction, as occasionally erring to increase its anthropomorphic and likeable qualities.
- In a future where AI is omnipresent, our species' self-worth may suffer an inescapable decline. Homo sapiens have forever been supreme, but because human evolution cannot outpace technological advancement, we will eventually be made ineffectual by these robots and wide-scale inferiority complexes may result.

If you have any questions about my paper, such as the limitations and future work, please email me:
danvieira21@gmail.com

6. References

Allan, S., & Gilbert, P. (1995). A social comparison scale: Psychometric properties and relationship to psychopathology. *Personality and Individual Differences*, *19*(3), 293-299.

Aronson, E., Willerman, B., & Floyd, J. (1966). The effect of a pratfall on increasing interpersonal attractiveness. *Psychonomic Science*, *4*(6), 227-228.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. International journal of social robotics, 1(1), 71-81.

Blanton, H., & Stapel, D. A. (2008). Unconscious and spontaneous and... complex: the three selves model of social comparison assimilation and contrast. *Journal of personality and social psychology*, *94*(6), 1018.

Botta, R. A. (1999). Television images and adolescent girls' body image disturbance. *Journal of communication*, *49*(2), 22-41.

Bruckenberger, U., Weiss, A., Mirnig, N., Strasser, E., Stadler, S., and Tscheligi, M. (2013). "The good, the bad, the weird: audience evaluation of a "real" robot in relation to science fiction and mass media," in *Proceedings of the International Conference on Social Robotics* (Bristol, UK: Springer), 301–310.

Buunk, B. P., Collins, R. L., Taylor, S. E., VanYperen, N. W., & Dakof, G. A. (1990). The affective consequences of social comparison: either direction has its ups and downs. *Journal of personality and social psychology*, *59*(6), 1238.

Cassell, J., & Thorisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, *13*(4-5), 519-538.

Chang, C. W., Lee, J. H., Po-Yao, C., Chin-Yeh, W., & Gwo-Dong, C. (2010). Exploring the possibility of using humanoid robots as instructional tools for teaching a second language in primary school. *Journal of Educational Technology & Society*, *13*(2), 13.

Clifford Nass, Jonathan Steuer, Ellen Tauber, and Heidi Reeder (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. In INTERACT'93 and CHI'93 conference companion on Human factors in computing systems, 111-112.

Collins, R. L. (2000). Among the better ones. *Handbook of social comparison: Theory and research*, 159-171.

Damiano, Luisa, and Paul Dumouchel. "Anthropomorphism in Human–Robot Co-evolution." *Frontiers in psychology* 9 (2018): 468.Festinger, L. (1954). A theory of social comparison processes. *Human relations*, *7*(2), 117-140.

Frey, C. B., & Osborne, M. A. (2017). The future of employment: how susceptible are jobs to computerisation?. *Technological Forecasting and Social Change*, *114*, 254-280.

Grace, K., Salvatier, J., Dafoe, A., Zhang, B., & Evans, O. (2017). When Will AI Exceed Human Performance? Evidence from AI Experts. *arXiv preprint arXiv:1705.08807*.

Gibbons, F.X. & Buunk, B.P. (1999). Individual differences in social comparison: The development of a scale of social comparison orientation. Journal of Personality and Social Psychology, 76, 129-142.

Gilbert, D. T., Giesler, R. B., & Morris, K. A. (1995). When comparisons arise. *Journal of personality and social psychology*, *69*(2), 227.

Kedia, G., Mussweiler, T., & Linden, D. E. (2014). Brain mechanisms of social comparison and their influence on the reward system. *NeuroReport*, *25*(16), 1255.

Krayer, A., Ingledew, D. K., & Iphofen, R. (2007). Social comparison and body image in adolescence: a grounded theory approach. *Health Education Research*, *23*(5), 892-903.

Lee, K. M., & Nass, C. (2005). Social-psychological origins of feelings of presence: Creating social presence with machine-generated voices. *Media Psychology*, *7*(1), 31-45.

Lockwood, P., & Kunda, Z. (1997). Superstars and me: Predicting the impact of role models on the self. Journal of Personality and Social Psychology, 73, 91–103.

Lohse, M. (2011). The role of expectations and situations in human-robot interaction. *New Front. Hum. Robot Interact.* 2, 35–56. doi:10.1075/ais.2.04loh

Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. *Frontiers in Robotics and AI*, *4*, 21.

Morse, S., & Gergen, K. J. (1970). Social comparison, self-consistency, and the concept of self. *Journal of personality and social psychology*, *16*(1), 148.

Nass, C., & Steuer, J. (1993). Voices, boxes, and sources of messages. *Human Communication Research*, *19*(4), 504-527.

Nass, C., & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of experimental psychology: applied*, *7*(3), 171.

Niu, S., McCrickard, D. S., & Harrison, S. (2015, June). Exploring humanoid factors of robots through transparent and reflective interactions. In *Collaboration Technologies and Systems (CTS), 2015 International Conference on* (pp. 47-54). IEEE.

Nomura, T., Suzuki, T., Kanda, T., & Kato, K. (2006). Altered attitudes of people toward robots: Investigation through the Negative Attitudes toward Robots Scale. In *Proc. AAAI-06 Workshop on Human Implications of Human-Robot Interaction* (Vol. 2006, pp. 29-35).

Prensky, M. (2001). Digital natives, digital immigrants part 1. *On the horizon*, *9*(5), 1-6.

Purington, A., Taft, J. G., Sannon, S., Bazarova, N. N., & Taylor, S. H. (2017, May). Alexa is my new BFF: social roles, user satisfaction, and personification of the amazon echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems* (pp. 2853-2859). ACM.

Pyszczynski, T., Greenberg, J., & LaPrelle, J. (1985). Social comparison after success and failure: Biased search for information consistent with a self-serving conclusion. *Journal of Experimental Social Psychology*, *21*(2), 195-211.

Ragni, M., Rudenko, A., Kuhnert, B., & Arras, K. O. (2016, August). Errare humanum est: Erroneous robots in human-robot interaction. In *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on* (pp. 501-506). IEEE.

Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, *5*(3), 313-323.

Stapel, D., & Koomen, W. (2000). Distinctness of others, mutability of selves: Their impact on self-evaluations. Journal of Personality and Social Psychology, 79, 1068–1087.

Suls, J., Martin, R., & Wheeler, L. (2002). Social comparison: Why, with whom, and with what effect?. *Current directions in psychological science*, *11*(5), 159-163.

Troop, N. A. (2016). Social Rank, Rank‐Related Life Events and Eating Pathology. *European Eating Disorders Review*, *24*(1), 75-77.

White, J. B., Langer, E. J., Yariv, L., & Welch, J. C. (2006). Frequent social comparisons and destructive emotions and behaviors: The dark side of social comparisons. *Journal of Adult Development*, *13*(1), 36-44.

Wang, F. Y., Zhang, J. J., Zheng, X., Wang, X., Yuan, Y., Dai, X., ... & Yang, L. (2016). Where does AlphaGo go: from church-turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, *3*(2), 113-120.

Wills, T.A. (1981). Downward comparison principles in social psychology. Psychological Bulletin, 90, 245–271.

Wood, J. V., Taylor, S. E., & Lichtman, R. R. (1985). Social comparison in adjustment to breast cancer. *Journal of personality and social psychology*, *49*(5), 1169.

Wood, J.V. (1989). Theory and research concerning social comparisons of personal attributes. Psychological Bulletin.