This is a preprint. The original article is published in HRI '22: Proceedings of the 2022 ACM/IEEE International Conference on Human-Robot Interaction. Please cite the published version at https://doi.org/10.1109/HRI53351.2022.9889520

Better than Us: The Role of Implicit Self-Theories in Determining Perceived Threat Responses in HRI

D. D. Allan
Human Interface Technology Lab NZ
University of Canterbury
Christchurch, New Zealand
0000-0001-9859-7033

Andrew J. Vonasch
Department of Psychology
University of Canterbury
Christchurch, New Zealand
0000-0002-2784-5420

Christoph Bartneck
Human Interface Technology Lab NZ
University of Canterbury
Christchurch, New Zealand
0000-0003-4566-4815

***Abstract*—Robots that are capable of outperforming human beings on mental and physical tasks provoke perceptions of threat. In this article we propose that implicit self-theory (core beliefs about the malleability of self-attributes, such as intelligence) is a determinant of whether one person experiences threat perception to a greater degree than another. We test for this possibility in a novel experiment in which participants watched a video of an apparently autonomous intelligent robot defeating human quiz players in a general knowledge game. Following the video, participants received either social comparison feedback, improvement-oriented feedback, or no feedback, and were then given the opportunity to play against the robot. We show that those who adopt a malleable self-theory (incremental theorists) are more likely to play against a robot after imagining losing to it, as well as exhibit more favorable responses and less identity threats than entity theorists (those adopting a fixed self-theory). Moreover, entity theorists (vs. incremental theorists) perceive autonomous intelligent robots to be significantly more threatening (both in terms of realistic and identity threats). These findings offer novel theoretical and practical implications, in addition to enriching the HRI literature by demonstrating that implicit self-theory is, in fact, an influential variable underpinning perceived threat.**

***Index Terms*—implicit self-theories, mindset, human–robot interaction, social robotics, identity threat, realistic threat, perception, robot acceptance**

I. Introduction

Perceived threat also known as "threat perception," is one of the more popular research themes to emerge within the field of HRI over the course of the last five years [e.g., 41, 112, 115, 74, 60, 54, 53]. The basic premise is as follows: machines that are increasingly humanlike and/or highly capable, simultaneously threaten human safety and resources (i.e., realistic threat), as well as human uniqueness and identity (i.e., identity threat). Several studies demonstrate the broad impact of perceived threat in HRI. [41] for example, found that high levels of similarity between social robots and humans, negatively impacted human group identity. The authors argued that similarity blurs category boundaries which, in consequence, undermines human uniqueness.

Yogeeswaran and colleagues [112] demonstrated, similarly, that a new generation of robots perceived to outperform humans on a variety of physical and mental tasks increased feelings of threat. This, according to the authors, is because human beings view such robots as members of a highly competent outgroup. Hence, these robots are considered to be plausible competition for resources and jobs, and, in addition, a threat to human identity and distinctiveness.

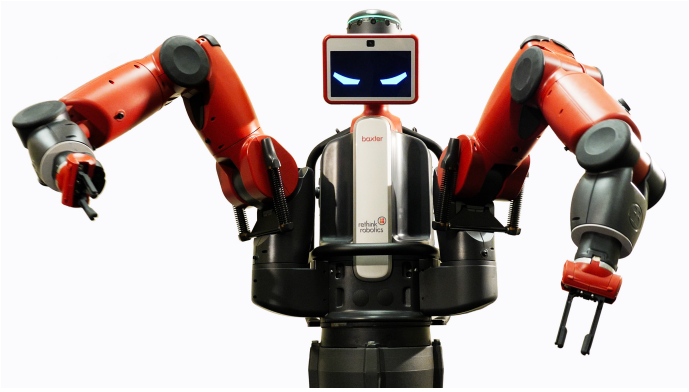

Fig. 1. The Baxter robot used in this experiment.

Correspondingly, in a study conducted by [115], participants who watched videos of purportedly autonomous robots (i.e., robots capable of disregarding human commands), considered robots to be significantly more threatening (both in terms of realistic and identity threats) than those who were shown a video of what appeared to be non-autonomous robots. What this finding indicates, the authors argue, is that when human beings experience a sense of perceived loss and control over robots, such robots are perceived as threatening.

Other avenues of research have largely buttressed these findings [e.g., 104, 53, 103, 54], particularly, Cha and his colleagues [15] have made a strong case that human-machine intellectual comparisons induce perceived threats. In one study, they showed that the defeat of human Go champion, Lee Sedol by Google's AlphaGo computer program threatened human distinctiveness [15, Study 1]. They also discovered that human intellectual comparison with machine intelligence adversely affects people's motivation towards intellectual tasks [15].

This finding is similar to that of [82], who found that participants who viewed the Google DeepMind Challenge (AlphaGo vs. Lee Sedol), experienced not only threats to human distinctiveness and resources, but increased feelings of helplessness.

The growing research interest in perceived threat has given rise to an important question: "what are the determinants of threat perception?" [94, p.630]. In attempting to answer this question, it seems, [115] has theorized that identity threat, in particular, "may, be universal and experienced by all people since it concerns humankind uniqueness and distinctiveness as a whole" [115, p.53].

In this article, we argue that this is not necessarily the case. More specifically, we suggest that perceptions of threat are not uniform for all people. Rather, we propose that an individual's implicit self-theory—underlying beliefs about whether self-attributes such as intellectual abilities are changeable or fixed [28, 32, 27]—could be an important determinant of perceived threat. We postulate that an incremental theory (those who believe in the malleability of self-attributes) decreases—whereas an entity theory (those who believe in the fixedness of self-attributes) increases—perceptions of threat. We test for this possibility in a novel experiment in which participants watched a video of an apparently intelligent robot defeating human opponents in a general knowledge quiz game. Following the video, participants received either social comparison feedback, improvement-oriented feedback, or no feedback; they were then given the opportunity to play against the robot.

Before elaborating further and explicating expected outcomes, we first summarize the relevant literature on implicit self-theory and provide the rationale for our hypotheses. We next report the findings of the abovementioned experiment. Finally, the theoretical and practical implications of our findings are discussed.

II. Theoretical framework and hypotheses

A. Implicit Self-Theory

As mentioned above, implicit self-theory (henceforth, IST), colloquially referred to as 'mindset,' pertains to one's underlying core beliefs regarding the plasticity of self-attributes, such as intelligence [35] and personality [29], among others [28]. ISTs exist on a continuum from the incremental theory, which assumes that a particular self-attribute is malleable and susceptible to change via effort, practice, and learning [32], to the entity theory, which assumes that such self-attributes are fixed and cannot be improved or changed [27].

A substantial and ever-increasing literature suggests that these opposing self-theories causally affect people's judgments [29], motivation [14], and behavior [28]. Whilst a comprehensive examination of the large IST literature is beyond the scope of this short article [but see 32, for review], six characteristics warrant specific mention. First, ISTs are distinct from familiar psychological theories such as the Big Five Model of personality [see 102], regulatory focus theory, and dual-process models [see 73], self-efficacy [see, 21], perceived control, and attribution style [see 97]. Second, the effects of ISTs appear to manifest most strikingly under conditions of difficulty, or in situations involving setbacks [110]. Third, the occurrence of each self-theory in a given population appears to be roughly equal [14, 75]. Fourth, while considered reasonably stable [35, 92] and able to be measured as a trait variable [66], ISTs can be (and indeed frequently are) temporarily manipulated, using scientific articles [10, 51] and other experimental stimuli [58]. Fifth, ISTs have been repeatedly shown to affect a vast array of self-attributes [27, 89] including, intelligence [11, 35, 92], personality [39, 18], morality [17, 55], 2017), and emotions [61, 105, 22], as well as downstream variables such as, evaluations of brands [83, 71], marketing messages [58, 113], technology [42, 101, 45], and social robots [3]. Finally, and perhaps not surprisingly given the above, ISTs appear to compel people to experience a theory-consistent view of reality, though the influence of which, typically occurs outside of awareness [27, 89]. Moving beyond these characteristics, we next summarize three specific interrelated lines of IST research that together constitute the basis for our hypotheses.

B. Implicit Self-Theory, Goals, and Responses to Challenges

One hallmark of ISTs is that they predispose people toward divergent goals [30, 37]. The incremental theory orients people towards learning goals, which involve effort (e.g., mastering a challenging task), and the potential for skill acquisition [34, 31]. As a consequence, incremental theorists tend to persist in the face of difficulty and attribute low performance to low effort [58, 75]. While an incremental theory is motivated toward learning goals, an entity theory is characterized by a fundamental propensity towards performance goals [36, 37], in which they seek to obtain approval and positive judgments from others [29] through displays of supposedly inherent and fixed pre-existing competencies [11], and the avoidance of failure [51], even if this requires cheating, giving up, withdrawing effort, or blaming others to hide potential inaptitude [30, 75]. For the entity theorist, the exertion of effort is viewed as a sign of incompetence [43], because if one has high competence, high effort is unnecessary [77]. Likewise, poor performance is considered to be a reflection of low ability or low intelligence [11, 52]. Nussbaum and Dweck [80] showed that entity and incremental theorists exhibit distinct patterns of response following failure. In one set of studies, college students who worked on a difficult task on which they inevitably failed, were given the option to examine the strategies of others who had performed either better, worse, or the same as they had (i.e., upward or downward social comparisons). Entity participants opted to view the strategies of those who had performed comparatively worse than they had. Incremental participants, however, viewed this as an opportunity to redress their performance by choosing to look at the strategies of people who had exceeded their performance [80, Studies 1 and 2]. A follow-up study confirmed that entity theorists felt better after defensively comparing themselves to poorer performers. Whereas incremental theorists felt better confronting and addressing their poor performance in an effort to improve on the task [80, Study 3]. This is consistent with previous findings which have shown that incremental theorists express greater optimism, resilience, and perseverance when responding to challenges and setbacks [e.g., 70, 52]. Entity theorists, by contrast, are sensitive to negative events [92, 37]

and respond with suboptimal stress mitigation strategies [25]. In the domain of HRI, Allan and colleagues [3] presented initial evidence that entity theorists, relative to incremental theorists, prefer a robot positioned as a personal servant (vs. a collaborative assistant). The reasons for this finding may stem in part from the entity theorists' desire to feel superior to others [27]. Consistent with this nascent literature, we would expect that in comparison to incremental theorists, entity theorists with their desire to not appear incompetent, will avoid playing against a winning intelligent robot. We would also expect entity theorists to evaluate a robot more favorably in the presence of social comparison feedback. Stated formally:

Hypothesis 1 (H1): Entity (vs. incremental) theorists will be less (vs. more) likely to play against the robot after imagining losing to it.

Hypothesis 2 (H2): Entity theorists will evaluate the robot in the social comparison condition more (vs. less) favorably than the robot in the improvement-oriented and control conditions.

C. Implicit Self-Theory and Responses to Feedback

A substantial stream of IST research has found that ISTs can arise from, and be reinforced by, significant others' (e.g., teachers, parents, and peers) feedback [46]. In one study [68], for example, university students with English as a second language failed a challenging English test and were given either ability-consoling feedback (e.g., "I'm sure you have great talent in other subjects."), improvement-oriented feedback (e.g., "If you put in the work, you'll be at the level of proficiency that you want."), or no feedback from a teacher. Students who received ability-consoling feedback reasoned that the teacher did not think they could improve and, in consequence, expressed an unwillingness to retake the test. Conversely, students who received the improvement-oriented feedback perceived the teacher to believe in their potential. According to Lou and Noels [68], the teacher's feedback strengthened and further increased students' endorsement of an incremental theory. Consistent with this notion, [4] recently found that ISTs and robot-delivered feedback (praise, in this case), interactively influenced people's evaluations of a social robot after a challenging task. In particular, they showed that entity theorists, compared to incremental theorists, evaluated a social robot as more intelligent and likable after it delivered ability praise (e.g., "You must be smart at these questions.") rather than effort praise (e.g., "You must have worked hard at these questions.") Moreover, incremental theorists, were unaffected by either praise type and evaluated the robot favorably irrespective of the praise it delivered. Together, these findings provide evidence of a matching effect between feedback and individual IST [see also 19, 47, 59, 88, 100]. Accordingly, we would expect incremental theorists, with their focus on developing ability, would evaluate a robot more favorably in the presence of feedback that emphasizes improvement. Stated formally:

Hypothesis 3 (H3): Incremental theorists will evaluate the robot in the improvement-oriented condition more (vs. less) favorably than the robot in the social comparison feedback and control conditions.

D. Implicit Self-Theory and Responses to Others

A well-established theme in the IST literature is that an individual's IST exerts a powerful influence on the attributions they make with respect to other people [34, 33, 38]. For instance, entity theorists have been found to make snap judgments about the traits of others [17], even in the presence of limited information [49]. In one set of studies, Yeager and colleagues [111], for example, found that adolescents with an entity theory interpreted ambiguous provocations from (unknown) peers to be hostile in intent, which in turn lead them to express relatively violent desires for revenge. Furthermore, individuals who hold entity theories, in comparison to those who hold incremental theories, are prone to making stereotypical judgments [67, 85, 86]. [67] found that entity theorists endorsed extreme judgments on global traits (i.e., bad-good, evil-virtuous) toward a range of target groups (e.g., ethnic, occupational, and even fictitious groups). Moreover, those endorsing an entity theory have been found to display a significant increase in self-esteem when promoting stereotypes [27]. Conversely, incremental theorists make fewer trait attributions [86, 66] and rather base their perceptions of people on situational and environmental factors [33, 34]. Of note is that stereotyping has long been identified as an integral component in the formation of outgroup attributions [see 81, for review]. Recall also from Section I that outgroup attributions have been linked to threat perception. Additionally, and perhaps most pertinent to our discussion, is a study in the domain of HRI, demonstrating that entity theorists, relative to incremental theorists, exhibit greater negative robot beliefs [3]. Notably, these findings fall in line with those of [115], which has connected negative robot beliefs to realistic and identity threats. Thus, we formulated the following hypotheses with regard to perceived threat:

Hypothesis 4 (H4): Incremental (vs. entity) theorists will rate robots in general as posing less (vs. more) identity relevant threats.

Hypothesis 5 (H5): Entity (vs. incremental) theorists will rate robots in general as posing more (vs. less) realistic threats.

III. Method

The experimental hypotheses and analysis plan were preregistered.[1] All data were collected during August 2021. The experimental protocol was reviewed and approved by the Human Research Ethics Committee of the University of Canterbury (HEC 2021/33).

A. Overview

We tested our hypotheses in an online human-subject experiment, which was administered through the Qualtrics online platform. Participants were drawn from Amazon Mechanical

[1] https://aspredicted.org/blind.php?x=/HY4_JJY.

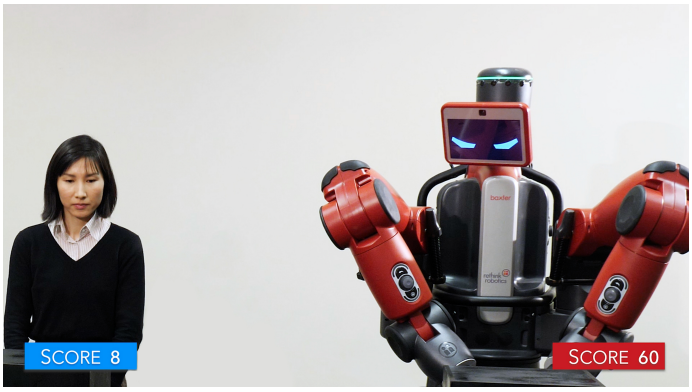

Fig. 2. An on-the-spot screenshot from the video showing the Baxter robot and a human player. Note. The individual pictured here has provided written informed consent to publish their image alongside the manuscript.

Turk (MTurk). While conducting online studies of this nature, has been well accepted [50] and increasingly recommended in HRI research [e.g., 40], here, our primary motivation was to access a large global and diverse sample[2] of participants [13].

The experiment was performed in four parts. First, we experimentally primed participants' ISTs, turning some toward an incremental theory and others toward an entity theory. The main justification for inducing ISTs, as opposed to measuring them as chronic orientations, was based on the findings of [3], who showed that MTurk workers who self-selected to take part in a study about robots, were prominently skewed in the direction of the incremental theory. This observation is not new and has been noted by other researchers from different domains [see 49, 56, 99, 57], suggesting that entity theorists might be less likely to participate in research studies than incremental theorists. Correspondingly, we sought to obtain a more theory-balanced sample.

In the second part of the experiment, we exposed participants to a video of people playing against (and losing to) an ostensibly autonomous intelligent robot in a general knowledge quiz game (see Fig. 2). It is worth mentioning here, that the use of video media featuring robots has been shown repeatedly to be effective in engendering people's experiences of, and attitudes towards, autonomous robots [e.g., 8], and most relevantly, perceived threats [see 112, 115].

In the third part of the experiment, participants were presented with either improvement-oriented, social comparison, or neutral feedback. As discussed in Section II-C, such feedback has consistently emerged as an interactive influence with ISTs on both people's responses to perceived failure [76], and judgments of others [68].

The fourth part of the experiment involved administering the dependent measures (described later in Section III-G).

[2]Of course, like most research conducted during a pandemic, this was also prompted by time and resource constraints, in combination with the myriad potential challenges associated with running physical experiments at this time [see 40].

B. Design

The experiment was a 2 (implicit self-theory: entity vs. incremental) $\times$3 (feedback: social comparison vs. improvement-oriented vs. control) between-subjects factorial design.

C. Recruitment

Participants were recruited through a task posted on the MTurk website. The posted task made clear that this was a study designed to survey people's "impressions of an intelligent autonomous game-playing robot." As well, prospective participants were informed that they would be required to (1) read a short paragraph describing "our approach to intelligent robots" and, (2) complete some questions associated with that paragraph. Furthermore, respondents were informed that they would be expected to watch a video of the robot and answer questions related to that video. This pretense afforded us the opportunity to unobtrusively administer the theory inductions, as well as providing a plausible context in which an intelligent robot could convincingly defeat human beings. Participation was contingent on (a) having completed >50 surveys with a minimal HIT approval rate of 98% or greater, (b) being located in either the USA, UK, Canada, Ireland, Australia, or New Zealand, and (c) having no experience with robotics and/or any expertise in AI (e.g., data science and machine learning). It should be noted here, that both technical expertise and experience with robotics, have been found consistently to predict favorable responses toward, and acceptance of, social robots [e.g., 23, 65, 26, 63, 107, 62, 96]. Thus, it seemed reasonable to exclude such individuals as they may have obscured our findings.

D. Participants

A total of 356 participants recruited through MTurk agreed to participate in exchange for $1.00. Fifty participants (14.08%) who either answered the manipulation check question incorrectly (e.g., "very nice," "good and ehw," "it was awesome"), failed the attention check, or who asked for their data to be excluded, were omitted from the dataset prior to analyses. The final sample consisted of 305 participants (173 male, 130 female, 2 with no gender reported), aged 18–29 (n=54), 30–39 (n=114), 40–49 (n=69), 50–59 (n=41), and 60 and over (n=27). Most participants (56.1%) ranked undergraduate education (some college education) as their highest level of educational attainment. While 26.2% held postgraduate degrees and 17.7% had a high school education alone. All participants provided consent prior to participation.

E. Procedure

Upon consent, participants were assigned randomly to read one of two articles, endorsing either an entity, or incremental theory of intelligence.[3] Afterwards, participants were asked to briefly summarize their respective article, and complete The Implicit Theories of Intelligence Scale [ITIS; 28], which

[3]The full manipulation articles are freely available at the Open Science Framework https://osf.io/t7f2j/.

served as a manipulation check[4] (described later in Section III-G). Participants were then instructed to watch a video of a general knowledge quiz match, in which human quiz players appeared to compete against an apparently intelligent and autonomous robot (described below in Section III-F). After viewing the video, participants were randomly assigned to one of three feedback conditions: improvement-oriented feedback, social comparison feedback, or neutral feedback (described below in Section III-F). Following the manipulation, participants completed the dependent measures (described in Section III-G), and provided their demographic details. Next, participants were debriefed about the true objective and the experimental rationale of the experiment, including the deception employed. Finally, participants were given an opportunity to withdraw their data, thanked, and compensated with $1.00.

F. Materials

1) Video: The video was designed to fit the tradition of real-world human-machine competitions, such as IBM Watson's Jeopardy! Challenge [16], and the Google DeepMind Challenge Match between Lee Sedol and AlphaGo [64]. The video lasted approximately seven and a half minutes and featured a Baxter Robot [44]. Baxter is a six foot three inch tall, humanoid robot, built initially for the manufacturing sector (see Fig. 1). Consistent with prior work [78], we designed custom eyes with a random eye blink, in addition to pre-programming the robot with scripted responses, prior to filming. A female American voice introduced the robot, and following [112] emphasized that Baxter had been "shown to outperform humans on both physical tasks, such as weight lifting and on mental tasks, such as chess and problem solving." Participants were then informed that what followed were "highlights of an intelligence challenge that took place between a Baxter Robot and players from a top university quiz team" (who were, in fact, actors performing scripted responses). Participants then watched as Baxter defeated the human players 4–0 in a best-of-four quiz game (see Fig. 3 for final scores). The quiz questions and answers were, in part, derived from the official YouTube channel of the British television quiz show The Chase.[5] The video can be viewed online.[6]

2) Feedback: On the basis of prior work in IST (discussed in Section II-C), we manipulated feedback by emphasizing two different feedback types: improvement-oriented, and social comparison, which have observed interactive matching effects with individual self-theories (i.e., improvement-oriented/incremental, social comparison/entity), After watching the video, participants were instructed to imagine that they had lost to the robot, in the same way the players in the video had. Participants then read that the robot had provided feedback. Following [68], all participants were told that they had not performed well: "You did not do well in the game." Subsequently, some received improvement-oriented feedback

[4]This priming procedure is based on prior IST research [see 98, 55, 48, 4].
[5]https://www.youtube.com/c/thechaseofficial/videos.
[6]https://youtu.be/VYPr-XUibFA.

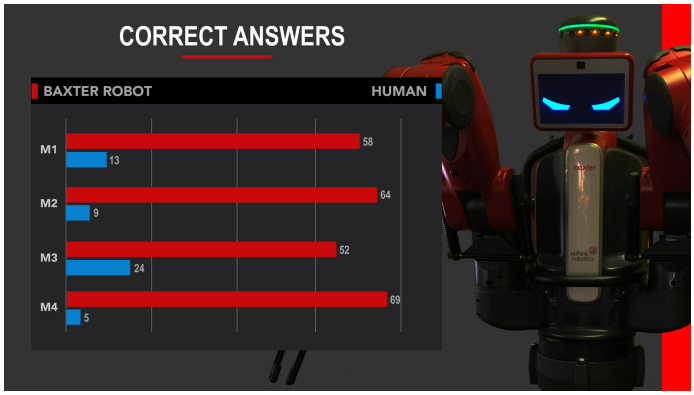

Fig. 3. A screenshot of the final scores of the game.

(e.g., "Like with many things, practice makes perfect. If you put in the work, you'll surely improve. Do you want to play me again?"). Whereas others were exposed to social comparison feedback (i.e., "Most humans are not naturally good at general knowledge. However, you did 37% better than the other humans. Do you want to play me again?"). It might be useful to point out that 37% was derived from [80], with the aim of provoking the entity theorists, into engaging in downward social comparisons (i.e., to think of the people in the video who had done relatively poorly), and thus defensively repair their self-esteem. Finally, the control condition received no additional feedback ("Do you want to play me again?"). The feedback scripts were adapted somewhat from [68] to suit our context.

G. Measures

1) Manipulation Check: The Implicit Theories of Intelligence Scale [ITIS; 28] was used as a manipulation check for participants' ISTs. This method has been used extensively elsewhere [see 66, 72, 91, 4]. The scale consists of 3 items that measure incremental beliefs (e.g., "You have a certain amount of intelligence, and you really can't do much to change it") and 3 items that measure entity theory beliefs (e.g., "You can always greatly change how intelligent you are"). Participants indicated agreement on a 6-point Likert scale (1=strongly disagree, 6=strongly agree). A final IST score was established by reverse scoring the incremental items and calculating a mean ITS score for all six items, with higher scores indicating a greater incremental view of intelligence. This measure demonstrated good internal consistency ($\alpha = 0.89, M = 3.60, SD = 1.22$).

2) Future Task Avoidance: A 5-item measure modified from [68] assessed participants' likelihood of playing against the robot after imagining losing to it (e.g., "I would try to avoid playing the robot again"). Participants indicated their agreement using a 5-point Likert scale (1=not at all, 5=very much). This measure demonstrated good internal consistency in our sample ($\alpha = 0.84, M = 3.23, SD = 1.08$).

3) Robot Evaluation: Robot Evaluation was assessed using the Robot Evaluation scale [3], a 3-item scale with variable 7-

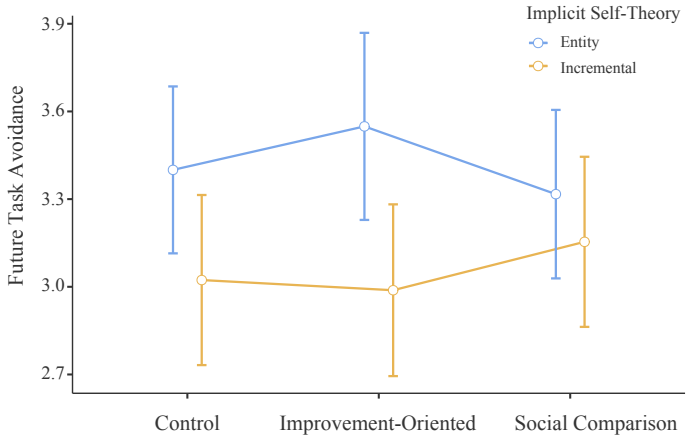

Fig. 4. Future Task Avoidance means and 95% confidence intervals for the effects of implicit self-theories and feedback.

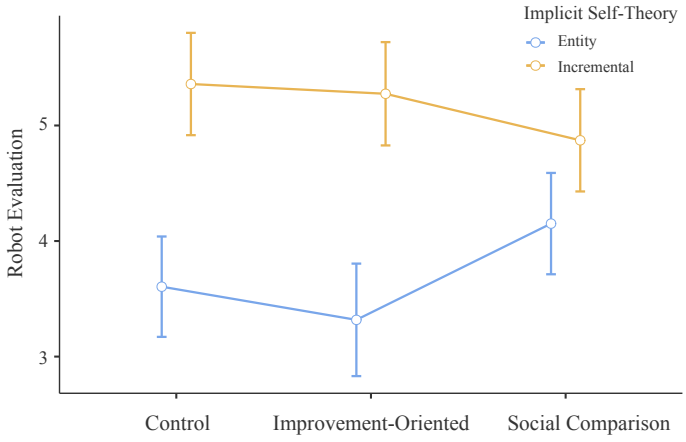

Fig. 5. Robot Evaluation means and 95% confidence intervals for the effects of implicit self-theories and feedback.

point Likert responses (1=very unfavorable/very bad/very negative, 7=very favorable/very good/very positive). The internal consistency of this measure was excellent ($\alpha = 0.94, M = 4.45, SD = 1.79$).

4) Identity Threat: Identity threat was assessed with the 5-item Identity Threat Measure [112]. Respondents indicated agreement with items (e.g., "Technological advancements in the area of robotics is threatening to human uniqueness") on a 7-point scale (1=strongly disagree, 7=strongly agree). This measure demonstrated excellent internal consistency in our sample ($\alpha = 0.93, M = 4.14, SD = 1.68$).

5) Realistic Threat: Realistic threat was assessed with the 5-item Realistic Threat Measure [112], Agreement with items (e.g., "in the long run, robots pose a direct threat to human safety and wellbeing") was indicated on a 7-point scale (1-strongly disagree, 7=strongly agree). This measure demonstrated good internal consistency ($\alpha = 0.80, M = 4.47, SD = 1.28$).

IV. Results

The dataset for this experiment is publicly available at the Open Science Framework.[7]

A. Main Analyses

1) Manipulation Check: Results of an independent sample t-test verified that participants who read the entity theory article reported significantly higher ($t(303) = 9.18, p < .001$) entity theory beliefs ($M = 3.03$) than those who read the incremental theory article ($M = 4.16$); thus, the manipulation of IST was successful. It is worth noting again, that lower scores on the ITIS represent higher entity beliefs.

2) Future Task Avoidance (H1): We used a 2 (implicit self-theory) $\times$3 (feedback) between-subjects ANOVA on future task performance. As expected, results showed a significant main effect for IST ($F(1, 299) = 8.97, p = 0.003$). We made no predictions regarding the effect of feedback condition, nor did we predict an interaction. No significant effect of feedback condition, ($F(2, 299) = 0.07, p = .931$), and no significant interaction emerged ($F(2, 299) = 0.86, p = .423$). Planned comparisons indicated that participants who were primed with entity theory demonstrated higher task avoidance ($M = 3.42, SD = 1.00$) than those primed with incremental theory ($M = 3.06, SD = 1.12, t(299) = 3.00, p = 0.003, d = 0.344, (95\%CI[0.116, 0.572]$). Consequently, H1 was supported (see Fig. 4).

3) Robot Evaluation (H2, H3): An additional 2 (implicit self-theory) $\times$3 (feedback) between-subjects ANOVA was performed to assess robot evaluation. Results revealed a significant main effect of IST ($F(1, 299) = 62.87, p < .001$), but no main effect of feedback condition, ($F(2, 299) = 0.50, p = .606$). However, a significant interaction between IST and feedback condition was observed ($F(2, 299) = 4.25, p = .015$). We subsequently conducted planned comparisons to examine the hypothesized differences between conditions. As expected, planned comparison results for entity theorists revealed that robot evaluation was significantly higher in the social comparison condition (M =4.15,SD=1.50) than in the improvement-oriented condition ($M = 3.32, SD = 1.93), t(299) = -2.50, p = 0.013, d = -.5136(95\%CI[-.9196, -.1076]$). However, robot evaluation was only marginally higher in the social comparison condition than in the control condition ($M = 3.6, SD = 1.83), t = -1.741, p = 0.083, d = -0.3366(95\%CI[-0.7181, 0.0449]$).

Planned comparison results for incremental theorists, revealed that improvement-oriented feedback ($M = 5.27, SD = 1.53$) led to a nonsignificantly more favorable robot evaluation than social comparison feedback ($M = 4.87, SD = 1.73), t(299) = 1.26, p = 0.209, d = 0.2483, (95\%CI[-0.1404, 0.6366]$). Furthermore, no significant difference in robot evaluation was found between the improvement-oriented condition and the control condition ($M = 5.50, SD = 1.15), t(299) = 0.264, p = 0.792, d = 0.0521(95\%CI[0.3358, 0.4399]$). Given these

[7]https://osf.io/rujge/.

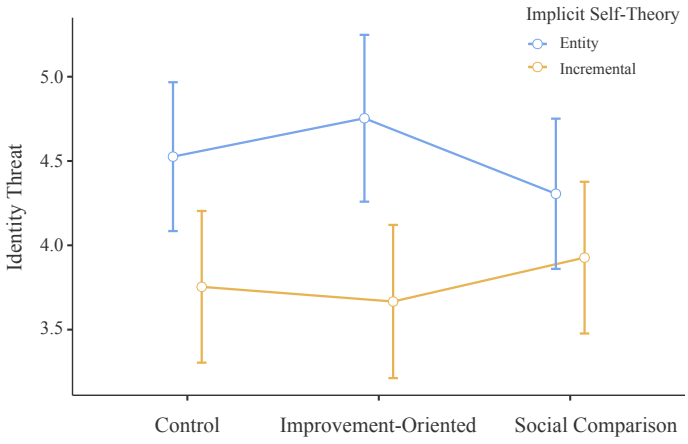

Fig. 6. Identity Threat means and 95% confidence intervals for the effects of implicit self-theories and feedback.

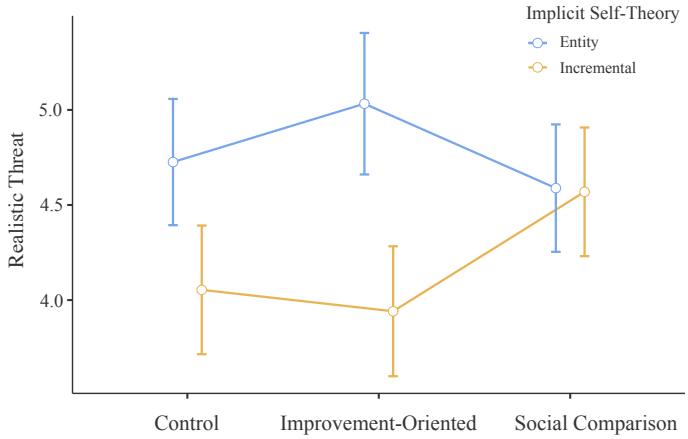

Fig. 7. Realistic Threat means and 95% confidence intervals for the effects of implicit self-theories and feedback.

results, H2 was partially supported, and H3 was not supported (see Fig. 5).

4) Identity Threat (H4): A 2 (implicit self-theory) $\times$3 (feedback) between-subjects ANOVA on identity threat was also conducted. As expected, a significant main effect was observed for IST ($F(1, 299) = 15.52, p < .001$). We made no predictions regarding a possible effect of feedback condition, nor did we predict an interaction. No significant effect of feedback condition, ($F(2, 299) = 0.09, p = .918$), and no significant interaction emerged ($F(2, 299) = 1.15, p = .318$). Planned comparisons revealed that incremental theorists perceived robots to be significantly less threatening to human identity and uniqueness ($M = 3.78, SD = 1.71$) than entity theorists ($M = 4.53, SD = 1.58), t(299) = 3.94, p < .001, d = 0.452(95\%CI[0.224, 0.681]$). This result supports H4 (see Fig. 6).

5) Realistic Threat (H5): We conducted a similar ANOVA test on realistic threat. As expected, a significant effect for IST, on realistic threat ($F(1, 299) = 17.428, p < .001$), was noted. We made no predictions regarding a possible effect of feedback condition, nor did we predict an interaction. No significant effect of feedback condition ($F(2, 299) = 0.61, p = .542$), was observed, however an unpredicted interaction between IST and feedback emerged ($F(2, 299) - = 4.74, p < .009$). Planned comparisons showed that entity theorists perceived robots to pose significantly more threat to employment, resources, and well-being ($M = 4.78, SD = 1.14$) than incremental theorists ($M = 4.19, SD = 1.35), t(299) = 4.17, p < .001, d = 0.479(95\%CI[0.250, 0.709]$). This finding supports H5 (see Fig. 7).

V. Discussion

In this research, we propose and find that a putatively autonomous and intelligent robot that defeats human beings in a general knowledge quiz game, does not evoke universal perceptions of threat (i.e., threaten human resources, jobs, and safety, as well as threaten human identity or distinctiveness) for every individual.

Specifically, we found that those embracing an incremental theory are more likely to play against an intelligent robot after imagining losing to it, thus supporting Hypothesis 1. Curiously however, those embracing an entity theory (who are thought to experience positive affect after comparing themselves to worse performers following failure), did not evaluate the robot significantly more favorably in the social comparison condition than in the control condition. As well, incremental theorists were not influenced by feedback, in so much as they evaluated the robot favorably regardless of the feedback they received. This finding, in particular, does not appear to be entirely consistent with the broader IST literature, linking improvement-oriented feedback to incremental theories [19, 47, 59, 88, 100] and the tendency to make positive judgments [i.e., 68], in consequence. However, this finding is consistent with that of [4] who recently showed that incremental theorists are unaffected by robot-delivered praise, and tend to respond more favorably to social robots in general [see also 3]. Consequently, Hypothesis 2 and Hypothesis 3 were not completely supported. That said, both findings were generally in the predicted direction, which might imply that a larger sample could confirm the expected effects. Consistent with Hypothesis 4, however, the results showed that compared to entity theorists, incremental theorists rated the robot as posing less identity relevant threats. We also observed that entity theorists relative to incremental theorists rated the robot as presenting more realistic threats, in support of Hypothesis 5.

Unexpectedly, an unpredicted interaction was suggested between incremental theory and social comparison feedback on realistic threat. Plausible reasons for this interaction are not evident, thereby necessitating further research to investigate this finding and to tease out possible explanations.

Most importantly however, these findings (particularly, the results of Hypotheses 4 and 5) expand the literature on perceived threat in HRI. As described earlier, this line of research finds that autonomous and intelligent robots and, more generally, "smart" machines incite threats to human

uniqueness, safety, and resources, which in turn, impedes positive perceptions of, and willingness to, engage with such technology [41, 112, 115, 74, 60, 54, 53]. We add to this literature by showing that IST is an important variable underpinning perceived threat. In so doing we extend our understanding of the influence of ISTs in HRI. Emerging work has begun to implicate ISTs as an important psychological factor in understanding aspects of HRI [3, 4]. Differences between incremental and entity theorists are apparent in the extent to which people experience robot anxiety [3], and robot delivered-praise, in addition to perceived robot intelligence and likeability [4].

Over and above their theoretical contributions, these findings may hold practical value for HRI designers and marketers. Our results would suggest that a robot presented as intelligent and capable of outperforming human beings could achieve greater acceptance and use among incremental theorists (vs. entity theorists). Whereas a robot that is presented as being less intelligent and unlikely to outperform human beings could increase the probability of acceptance among entity theorists, for example.

A. Limitations and future work

The contribution of our work is qualified by certain limitations, which could be addressed in future research. First, generalizations from our findings are obscured by the robot that was used. In other words, a Baxter Robot, which as a six foot three inch humanoid robot, cuts an imposing figure (see Fig. 1). Thus, the obvious question is whether the results reported in this article, would be observed with different robot types. Future work using alternative aesthetic robot forms (e.g., mechanomorphic, android), can shed light on this question.

Crucially, more research is needed to determine whether the effects noted here emerge following a real-life human-robot interaction. Although recent work has found fairly strong correlations of findings between online and in-person experiments [5], the imagining of robot delivered feedback following a perceived loss, is hardly the same thing as receiving direct feedback from a robot [50] in a real-life encounter (especially after competing against and losing to it). Accordingly, as physical experiments become feasible again, it is important that prospective research uses an in-person experimental design.

Future studies might also include a human control condition comparable to the robot condition in terms of general knowledge and perceived intelligence (e.g., a professional quizzer). Another limitation, although not within the scope of our hypotheses, might be that we did not assess well-known covariate influences that may interact or confound the interpretation of data [e.g., 3, 103, 109]. It is important for future studies to examine covariate influences to better understand the significance of our findings.

Our decision to use the MTurk platform might also have impeded our findings [114, 2]. While past research indicates that Mturk samples provide valid and reliable data [e.g., 9, 13], a recent stream of studies have reported many troubling trends regarding samples drawn from MTurk [see 108, 84]. These include, but are not limited to, participant carelessness [90], insufficient effort [69], and fraudulent responding [1]. Hence, although we took measures to strengthen the integrity of our data, such as placing attention and validity checks in the survey [93, 20], it is still a possibility that some respondents may have provided dishonest or poor quality responses [6, 24]. Relatedly, the measures used were obtained via self-report, which although theoretically substantiated and practically applicable, may be vulnerable to bias [e.g., participants may provide self-enhancing responses; 87].

One other limitation is that participants were excluded if they had experience with robotics and AI, and/or did not reside in an English-speaking country. This exclusion criterion limits the external validity of our findings. Therefore, additional studies employing alternative samples, different eligibility criteria, and more objective measures [e.g., use of an electroencephalogram; 12] seems both warranted and potentially beneficial.

Despite these limitations, the present findings open up avenues for future work. One interesting area to be explored is whether ISTs promote or reduce individuals' proclivity to exhibit aggressive behaviors towards robots [7, 79]. Our findings imply that an entity theory precipitates greater identity and realistic threats in response to intelligent capable robots. As noted, entity beliefs have been associated with greater negative attitudes regarding robots [3], and shown to provoke aggressive reactions to ambiguous provocations [111]. It seems reasonable to propose, then, that entity theorists may be prone to acts of robot-directed aggression [95]. Contrastingly, incremental theorists who tend to respond more positively to robots [3, 4] may be more likely to intervene when a robot is being mistreated or abused [106]. Of course, future research is needed to empirically test these propositions.

VI. Conclusion

The present findings indicate that significant individual variability exists with respect to perceived threat, among those exposed to an apparently intelligent quiz-playing robot (that defeats human beings). We show that some people (incremental theorists) are rather receptive to robots, that can both physically and mentally outperform human beings. In contrast, others (entity theorists) view such robots as threatening and thus rate them unfavorably. These findings support emerging research indicating that IST serves as an important variable in determining how human beings respond to robots.

Acknowledgments

The authors would like to thank Jonathan Wiltshire for providing diligent and reliable technical support during this project. The authors would also like to thank Rob Stowell and the University of Canterbury Video Production team, who helped with filming—on relatively short notice. Thanks is due, finally, to the anonymous reviewers and the HRI 2022 Program Committee for their helpful comments in revising the manuscript. This work was supported by the HIT Lab NZ and the University of Canterbury, New Zealand.

REFERENCES

[1] Jon Agley, Yunyu Xiao, Rachael Nolan, and Lilian Golzarri-Arroyo. Quality control questions on amazon's mechanical turk (mturk): A randomized trial of impact on the usaudit, phq-9, and gad-7. *Behavior Research Methods*, 2021. ISSN 1554-3528. URL https://doi.org/10.3758/s13428-021-01665-8.

[2] Herman Aguinis, Isabel Villamor, and Ravi S Ramani. Mturk research: Review and recommendations. *Journal of Management*, 47(4):823–837, 2021. ISSN 0149-2063. URL https://doi.org/10.1177/0149206320969787.

[3] D. D. Allan, Andrew J. Vonasch, and Christoph Bartneck. The doors of social robot perception: The influence of implicit self-theories. *International Journal of Social Robotics*, 14(1):127–140, 2022. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-021-00767-9.

[4] D. D. Allan, Andrew J. Vonasch, and Christoph Bartneck. "i have to praise you like i should?" the effects of implicit self-theories and robot-delivered praise on evaluations of a social robot. *International Journal of Social Robotics*, 2022. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-021-00848-9.

[5] Franziska Babel, Johannes Kraus, Philipp Hock, Hannah Asenbauer, and Martin Baumann. Investigating the validity of online robot evaluations: Comparison of findings from an one-sample online and laboratory study. In *Companion of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, page 116–120, 2021. URL https://doi.org/10.1145/3434074.3447141.

[6] Ard J Barends and Reinout E de Vries. Noncompliant responding: Comparing exclusion criteria in mturk personality research to improve data quality. *Personality and individual differences*, 143:84–89, 2019. ISSN 0191-8869. URL https://doi.org/10.1016/j.paid.2019.02.015.

[7] Christoph Bartneck and Jun Hu. Exploring the abuse of robots. *Interaction Studies*, 9(3):415–433, 2008. URL https://doi.org/10.1075/is.9.3.04bar.

[8] Christoph Bartneck and Merel Keijsers. The morality of abusing a robot. *Paladyn, Journal of Behavioral Robotics*, 11(1):271–283, 2020. ISSN 2081-4836. URL https://doi.org/10.1515/pjbr-2020-0017.

[9] Christoph Bartneck, Andreas Duenser, Elena Moltchanova, and Karolina Zawieska. Comparing the similarity of responses received from studies in amazon's mechanical turk to studies conducted online and with direct recruitment. *PLoS One*, 10(4):e0121595, 2015. ISSN 1932-6203. URL https://doi.org/10.1371/journal.pone.0121595.

[10] Randall Scott Bergen. *Beliefs about intelligence and achievement-related behaviors*. Unpublished doctoral thesis, University of Illinois at Urbana-Champaign, 1991.

[11] Lisa S Blackwell, Kali H Trzesniewski, and Carol Sorich Dweck. Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child development*, 78(1):246–263, 2007. ISSN 0009-3920. URL https://doi.org/10.1111/j.1467-8624.2007.00995.x.

[12] Francesco Bossi, Cesco Willemse, Jacopo Cavazza, Serena Marchesi, Vittorio Murino, and Agnieszka Wykowska. The human brain reveals resting state activity patterns that are predictive of biases in attitudes toward robots. *Science robotics*, 5(46), 2020. ISSN 2470-9476. URL https://doi.org/10.1126/scirobotics.abb6652.

[13] Michael Buhrmester, Tracy Kwang, and Samuel D. Gosling. *Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data?* Methodological issues and strategies in clinical research, 4th ed. American Psychological Association, Washington, DC, US, 2016. ISBN 1-4338-2091-9. URL https://doi.org/10.1037/14805-009.

[14] Jeni L Burnette, Ernest H O'Boyle, Eric M VanEpps, Jeffrey M Pollack, and Eli J Finkel. Mind-sets matter: A meta-analytic review of implicit theories and self-regulation. *Psychological bulletin*, 139(3):655, 2013. ISSN 1939-1455. URL https://doi.org/10.1037/a0029531.

[15] Youngjae Cha, Sojung Baek, Grace Ahn, Hyoungsuk Lee, Boyun Lee, Ji-eun Shin, and Dayk Jang. Compensating for the loss of human distinctiveness: The use of social creativity under human–machine comparisons. *Computers in Human Behavior*, 103:80–90, 2020. ISSN 0747-5632. URL https://doi.org/10.1016/j.chb.2019.08.027.

[16] Raman Chandrasekar. Elementary? question answering, ibm's watson, and the jeopardy! challenge. *Resonance*, 19 (3):222–241, 2014. ISSN 0971-8044. URL https://doi.org/10.1007/s12045-014-0029-7.

[17] Chi-yue Chiu, Carol S Dweck, Jennifer Yuk-yue Tong, and Jeanne Ho-ying Fu. Implicit theories and conceptions of morality. *Journal of Personality and Social Psychology*, 73(5):923, 1997. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.73.5.923.

[18] Chi-yue Chiu, Ying-yi Hong, and Carol S Dweck. Lay dispositionism and implicit theories of personality. *Journal of personality and social psychology*, 73(1):19, 1997. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.73.1.19.

[19] Andrei Cimpian, Holly-Marie C Arce, Ellen M Markman, and Carol S Dweck. Subtle linguistic cues affect children's motivation. *Psychological Science*, 18(4):314–316, 2007. ISSN 0956-7976. URL https://doi.org/10.1111/j.1467-9280.2007.01896.x.

[20] Cihan Cobanoglu, Muhittin Cavusoglu, and Gozde Turktarhan. A beginner's guide and best practices for using crowdsourcing platforms for survey research: The case of amazon mechanical turk (mturk). *Journal of Global Business Insights*, 6(1):92–97, 2021. ISSN 2640-6470. URL https://doi.org/10.5038/2640-6489.6.1.1177.

[21] Krista De Castella and Donald Byrne. My intelligence may be more malleable than yours: the revised implicit theories of intelligence (self-theory) scale is a better predictor of achievement, motivation, and student disengagement. *European Journal of Psychology of Education*, 30(3):245–267, 2015. ISSN 1878-5174. URL https://doi.org/10.1007/s10212-015-0244-y.

[22] Krista De Castella, Philippe Goldin, Hooria Jazaieri, Michal Ziv, Carol S Dweck, and James J Gross. Beliefs about emotion: Links to emotion regulation, well-being, and psychological distress. *Basic and Applied Social Psychology*, 35(6):497–505, 2013. ISSN 0197-3533. URL https://doi.org/10.1080/01973533.2013.840632.

[23] Maartje MA De Graaf and Somaya Ben Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and autonomous systems*, 61(12):1476–1486, 2013. ISSN 0921-8890. URL https://doi.org/10.1016/j.robot.2013.07.007.

[24] Sean A Dennis, Brian M Goodson, and Christopher A Pearson. Online worker fraud and evolving threats to the integrity of mturk data: A discussion of virtual private servers and the limitations of ip-based screening procedures. *Behavioral Research in Accounting*, 32(1):119–134, 2020. ISSN 1558-8009. URL http://dx.doi.org/10.2139/ssrn.3233954.

[25] Julie Doron, Yannick Stephan, Julie Boiché, and Christine Le Scanff. Coping with examinations: Exploring relationships between students' coping strategies, implicit theories of ability, and perceived control. *British Journal of Educational Psychology*, 79(3):515–528, 2009. ISSN 0007-0998. URL http://doi.org/10.1348/978185409x402580.

[26] Melanie Dudek, Stefanie Baisch, Monika Knopf, and Thorsten Kolling. "this isn't me!": The role of age-related self- and user images for robot acceptance by elders. *International Journal of Social Robotics*, 2020. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-020-00678-1.

[27] Carol Dweck. *Mindset-updated edition: Changing the way you think to fulfil your potential*. Hachette UK, 2017. ISBN 1472139968.

[28] Carol S Dweck. *Self-Theories: Their Role in Motivation, Personality, and Development. Essays in Social Psychology*. Psychology Press, 2000. ISBN 9781841690247. URL https://doi.org/10.4324/9781315783048.

[29] Carol S Dweck. *Self-theories: Their role in motivation, personality, and development*. Psychology press, 2013. ISBN 1317710339.

[30] Carol S Dweck and Ellen L Leggett. A social-cognitive approach to motivation and personality. *Psychological review*, 95(2):256, 1988. ISSN 1939-1471. URL https://doi.org/10.1037/0033-295X.95.2.256.

[31] Carol S Dweck and Lisa Sorich. Mastery-oriented thinking. *Coping*, 11:232–251, 1999. URL https://doi.org/10.1093/med:psych/9780195119343.003.0011.

[32] Carol S. Dweck and David S. Yeager. Mindsets: A view from two eras. *Perspectives on Psychological Science*, 14 (3):481–496, 2019. URL https://journals.sagepub.com/doi/abs/10.1177/1745691618804166.

[33] Carol S Dweck, Ying yi Hong, and Chi yue Chiu. Implicit theories individual differences in the likelihood and meaning of dispositional inference. *Personality and Social Psychology Bulletin*, 19(5):644–656, 1993. ISSN 0146-1672. URL https://doi.org/10.1177/0146167293195015.

[34] Carol S Dweck, Chi-yue Chiu, and Ying-yi Hong. Implicit theories and their role in judgments and reactions: A word from two perspectives. *Psychological inquiry*, 6(4):267–285, 1995. ISSN 1047-840X. URL https://doi.org/10.1207/s15327965pli0604_1.

[35] Carol S Dweck, Chi-yue Chiu, and Ying-yi Hong. Implicit theories: Elaboration and extension of the model. *Psychological inquiry*, 6(4):322–333, 1995. ISSN 1047-840X. URL https://doi.org/10.1207/s15327965pli0604_12.

[36] CS Dweck and J Bempechat. Theories of intelligence and achievement motivation. *Learning and motivation in the classroom. Hillsdale, NJ: Erlbaum*, 1983.

[37] E. S. Elliott and C. S. Dweck. Goals: an approach to motivation and achievement. *J Pers Soc Psychol*, 54(1):5–12, 1988. ISSN 1939-1315. URL https://doi.org/10.1037//0022-3514.54.1.5.

[38] Cynthia A Erdley and Carol S Dweck. Children's implicit personality theories as predictors of their social judgments. *Child development*, 64(3):863–878, 1993. ISSN 0009-3920. URL https://doi.org/10.2307/1131223.

[39] Cynthia A Erdley, Catherine C Loomis, Kathleen M Cain, and Frances Dumas-Hines. Relations among children's social goals, implicit personality theories, and responses to social failure. *Developmental psychology*, 33(2):263, 1997. ISSN 1939-0599. URL https://doi.org/10.1037/0012-1649.33.2.263.

[40] David Feil-Seifer, Kerstin S Haring, Silvia Rossi, Alan R Wagner, and Tom Williams. Where to next? the impact of covid-19 on human-robot interaction research. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1): 1–7, 2020. ISSN 2573-9522. URL https://doi.org/10.1145/3405450.

[41] Francesco Ferrari, Maria Paola Paladino, and Jolanda Jetten. Blurring human–machine distinctions: anthropomorphic appearance in social robots as a threat to human distinctiveness. *International Journal of Social Robotics*, 8(2): 287–302, 2016. ISSN 1875-4791. URL https://doi.org/10.1007/s12369-016-0338-y.

[42] Lawrence Hoc Nang Fong, Irene Cheng Chu Chan, Rob Law, and Tuan Phong Ly. *The Mechanism that Links the Implicit Theories of Intelligence and Continuance of Information Technology: Evidence from the Use of Mobile Apps to Make Hotel Reservations*, pages 323–335. Springer, 2018. URL https://doi.org/10.1007/978-3-319-72923-7_25.

[43] Heidi Grant and Carol S Dweck. Clarifying achievement goals and their impact. *Journal of personality and social psychology*, 85(3):541, 2003. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.85.3.541.

[44] Erico Guizzo and Evan Ackerman. How rethink robotics built its new baxter robot worker. *IEEE spectrum*, 7, 2012.

[45] Ajmal Hafeez. *Promoting upskilling: How a situational growth mindset increases consumers' adoption of really new products*. Unpublished doctoral dissertation, University of South-Eastern Norway, 2019.

[46] Kyla Haimovitz and Carol S Dweck. The origins of children's growth and fixed mindsets: New research and a new proposal. *Child Development*, 88(6):1849–1859, 2017. ISSN 0009-3920. URL https://doi.org/10.1111/cdev.12955.

[47] Kyla Haimovitz and Jennifer Henderlong Corpus. Effects of person versus process praise on student motivation: Stability and change in emerging adulthood. *Educational Psychology*, 31(5):595–609, 2011. ISSN 0144-3410. URL https://doi.org/10.1080/01443410.2011.585950.

[48] Bing Han, Liangyan Wang, and Xiang Li. To collaborate or serve? effects of anthropomorphized brand roles and implicit theories on consumer responses. *Cornell Hospitality Quarterly*, 61(1):53–67, 2019. ISSN 1938-9655. URL https://doi.org/10.1177/1938965519874879.

[49] Peter A Heslin, Gary P Latham, and Don VandeWalle. The effect of implicit person theory on performance appraisals. *Journal of Applied Psychology*, 90(5):842, 2005. ISSN 1939-1854. URL https://doi.org/10.1037/0021-9010.90.5.842.

[50] Guy Hoffman and Xuan Zhao. A primer for conducting experiments in human–robot interaction. *J. Hum.-Robot Interact.*, 10(1):Article 6, 2020. URL https://doi.org/10.1145/3412374.

[51] Ying-yi Hong, Chi-yue Chiu, Carol S Dweck, and Russell Sacks. Implicit theories and evaluative processes in person cognition. *Journal of Experimental Social Psychology*, 33(3):296–323, 1997. ISSN 0022-1031. URL https://doi.org/10.1006/jesp.1996.1324.

[52] Ying-yi Hong, Chi-yue Chiu, Carol S Dweck, Derrick M-S Lin, and Wendy Wan. Implicit theories, attributions, and coping: a meaning system approach. *Journal of Personality and Social psychology*, 77(3):588, 1999. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.77.3.588.

[53] Quirien RM Hover, Ella Velner, Thomas Beelen, Mieke Boon, and Khiet P Truong. Uncanny, sexy, and threatening robots: The online community's attitude to and perceptions of robots varying in humanlikeness and gender. In *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction*, pages 119–128, 2021. URL https://doi.org/10.1145/3434073.3444661.

[54] Hsien-Long Huang, Li-Keng Cheng, Pi-Chuan Sun, and Szu-Jung Chou. The effects of perceived identity threat and realistic threat on the negative attitudes and usage intentions toward hotel service robots: The moderating effect of the robot's anthropomorphism. *International Journal of Social Robotics*, 2021. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-021-00752-2.

[55] Niwen Huang, Shijiang Zuo, Fang Wang, Pan Cai, and Fengxiang Wang. The dark side of malleability: Incremental theory promotes immoral behaviors. *Frontiers in psychology*, 8:1341, 2017. ISSN 1664-1078. URL https://doi.org/10.3389/fpsyg.2017.01341.

[56] Jamie S Hughes. Support for the domain specificity of implicit beliefs about persons, intelligence, and morality. *Personality and Individual Differences*, 86:195–203, 2015. ISSN 0191-8869. URL https://doi.org/10.1016/j.paid.2015.05.042.

[57] Tatiana Iwai and João Vinícius de França Carvalho. Can a leopard change its spots? the effects of implicit theories of personality on forgiveness via attributions of behavioral stability. *Personality and Individual Differences*, 157: 109800, 2020. ISSN 0191-8869. URL https://doi.org/10.1016/j.paid.2019.109800.

[58] Shailendra Pratap Jain, Pragya Mathur, and Durairaj Maheswaran. The influence of consumers' lay theories on approach/avoidance motivation. *Journal of Marketing Research*, 46(1):56–65, 2009. ISSN 0022-2437. URL https://doi.org/10.1509/jmkr.46.1.56.

[59] Melissa L Kamins and Carol S Dweck. Person versus process praise and criticism: Implications for contingent self-worth and coping. *Developmental psychology*, 35(3):835, 1999. ISSN 1939-0599. URL https://doi.org/10.1037/0012-1649.35.3.835.

[60] Kimon Kieslich, Marco Lünich, and Frank Marcinkowski. The threats of artificial intelligence scale (tai). *International Journal of Social Robotics*, 2021. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-020-00734-w.

[61] Ronnel B King and Elmer D dela Rosa. Are your emotions under your control or not? implicit theories of emotion predict well-being via cognitive reappraisal. *Personality and Individual Differences*, 138:177–182, 2019. ISSN 0191-8869. URL https://doi.org/10.1016/j.paid.2018.09.040.

[62] Barbara Kühnlenz, Kolja Kühnlenz, Fabian Busse, Pascal Förtsch, and Maximilian Wolf. Effect of explicit emotional adaptation on prosocial behavior of humans towards robots depends on prior robot experience. In *2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 275–281. IEEE, 2018. ISBN 1538679809. URL https://doi.org/10.1109/ROMAN.2018.8525515.

[63] Rita Latikka, Nina Savela, Aki Koivula, and Atte Oksanen. Attitudes toward robots as equipment and coworkers and the impact of robot autonomy level. *International Journal of Social Robotics*, 2021. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-020-00743-9.

[64] Chang-Shing Lee, Mei-Hui Wang, Shi-Jim Yen, Ting-Han Wei, I-Chen Wu, Ping-Chiang Chou, Chun-Hsun Chou, Ming-Wan Wang, and Tai-Hsiung Yan. Human vs. computer go: Review and prospect [discussion forum]. *IEEE Computational Intelligence Magazine*, 11(3):67–72, 2016. URL https://doi.org/10.1109/MCI.2016.2572559.

[65] Hee Rin Lee and Selma Šabanović. Culturally variable preferences for robot design and use in south korea, turkey, and the united states. In *2014 9th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 17–24. IEEE, 2014. ISBN 1450326587. URL https://doi.org/10.1145/2559636.2559676.

[66] Sheri R Levy and Carol S Dweck. Trait-versus process-focused social judgment. *Social Cognition*, 16(1):151–172, 1998. ISSN 0278-016X. URL https://doi.org/10.1521/soco.1998.16.1.151.

[67] Sheri R Levy, Steven J Stroessner, and Carol S Dweck. Stereotype formation and endorsement: The role of implicit theories. *Journal of Personality and Social Psychology*, 74(6):1421, 1998. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.74.6.1421.

[68] Nigel Mantou Lou and Kimberly Ann Noels. "does my teacher believe i can improve?": The role of meta-lay theories in esl learners' mindsets and need satisfaction. *Frontiers in psychology*, 11:1417, 2020. ISSN 1664-1078. URL https://doi.org/10.3389/fpsyg.2020.01417.

[69] Lu Lu, Nathan Neale, Nathaniel D Line, and Mark Bonn. Improving data quality using amazon mechanical turk through platform setup. *Cornell Hospitality Quarterly*, 2021. ISSN 1938-9655. URL https://doi.org/10.1177/1938965521025475.

[70] Jennifer A Mangels, Brady Butterfield, Justin Lamb, Catherine Good, and Carol S Dweck. Why do beliefs about intelligence influence learning success? a social cognitive neuroscience model. *Social cognitive and affective neuroscience*, 1(2):75–86, 2006. ISSN 1749-5016. URL https://doi.org/10.1093/scan/nsl013.

[71] Pragya Mathur, Shailendra P Jain, and Durairaj Maheswaran. Consumers' implicit theories about personality influence their brand personality judgments. *Journal of Consumer Psychology*, 22(4):545–557, 2012. ISSN 1057-7408. URL https://doi.org/10.1016/j.jcps.2012.01.005.

[72] Pragya Mathur, Lauren Block, and Ozge Yucel-Aybat. The effects of goal progress cues: An implicit theory perspective. *Journal of Consumer Psychology*, 24(4):484–496, 2014. ISSN 1057-7408. URL https://doi.org/10.1016/j.jcps.2014.03.003.

[73] Pragya Mathur, HaeEun Helen Chun, and Durairaj Maheswaran. Consumer mindsets and self-enhancement: Signaling versus learning. *Journal of Consumer Psychology*, 26(1):142–152, 2016. ISSN 1057-7408. URL http://www.sciencedirect.com/science/article/pii/S1057740815000674.

[74] Antonia Meissner, Angelika Trübswetter, Antonia S Conti-Kufner, and Jonas Schmidtler. Friend or foe? understanding assembly workers' acceptance of human-robot collaboration. *ACM Transactions on Human-Robot Interaction (THRI)*, 10(1):1–30, 2020. ISSN 2573-9522. URL https://doi.org/10.1145/3399433.

[75] Daniel C Molden and Carol S Dweck. Finding" meaning" in psychology: a lay theories approach to self-regulation, social perception, and social development. *American Psychologist*, 61(3):192, 2006. ISSN 1935-990X. URL https://doi.org/10.1037/0003-066X.61.3.192.

[76] Claudia M Mueller and Carol S Dweck. Praise for intelligence can undermine children's motivation and performance. *Journal of personality and social psychology*, 75(1):33, 1998. ISSN 1939-1315. URL https://doi.org/10.1037//0022-3514.75.1.33.

[77] Mary C Murphy and Carol S Dweck. Mindsets shape consumer behavior. *Journal of Consumer Psychology*, 26(1): 127–136, 2016. ISSN 1057-7408. URL https://doi.org/10.1016/j.jcps.2015.06.005.

[78] Kyle Nash, Johanna M Lea, Thomas Davies, and Kumar Yogeeswaran. The bionic blues: Robot rejection lowers self-esteem. *Computers in human behavior*, 78:59–63, 2018. ISSN 0747-5632. URL https://doi.org/10.1016/j.chb.2017.09.018.

[79] Tatsuya Nomura, Takayuki Kanda, Hiroyoshi Kidokoro, Yoshitaka Suehiro, and Sachie Yamada. Why do children abuse robots? *Interaction Studies*, 17(3):347–369, 2016. ISSN 1572-0373. URL https://doi.org/10.1145/2701973.2701977.

[80] A David Nussbaum and Carol S Dweck. Defensiveness versus remediation: Self-theories and modes of self-esteem maintenance. *Personality and Social Psychology Bulletin*, 34(5):599–612, 2008. ISSN 0146-1672. URL https://doi.org/10.1177/0146167207312960.

[81] Penelope J Oakes, S Alexander Haslam, and John C Turner. *Stereotyping and social reality*. Blackwell Publishing, 1994. ISBN 0631188711.

[82] Changhoon Oh, Taeyoung Lee, Yoojung Kim, SoHyun Park, Saebom Kwon, and Bongwon Suh. Us vs. them: Understanding artificial intelligence technophobia over the google deepmind challenge match. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pages 2523–2534, New York, NY, USA, 2017. ACM. ISBN 978-1-4503-4655-9. URL 10.1145/3025453.3025539.

[83] Ji Kyung Park and Deborah Roedder John. Got to get you into my life: Do brand personalities rub off on consumers? *Journal of consumer research*, 37(4):655–669, 2010. ISSN 1537-5277. URL https://doi.org/10.1086/655807.

[84] Eyal Peer, David M Rothschild, Zak Evernden, Andrew Gordon, and Ekaterina Damer. Mturk, prolific or panels? choosing the right audience for online research. *Choosing the right audience for online research (January 10, 2021)*, 2021. URL http://dx.doi.org/10.2139/ssrn.3765448.

[85] Jason E Plaks, Steven J Stroessner, Carol S Dweck, and Jeffrey W Sherman. Person theories and attention allocation: Preferences for stereotypic versus counterstereotypic information. *Journal of personality and social psychology*, 80 (6):876, 2001. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.80.6.876.

[86] Jason E Plaks, Heidi Grant, and Carol S Dweck. Violations of implicit theories and the sense of prediction and control: implications for motivated person perception. *Journal of Personality and Social Psychology*, 88(2):245, 2005. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.88.2.245.

[87] Philip M Podsakoff, Scott B MacKenzie, Jeong-Yeon Lee, and Nathan P Podsakoff. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of applied psychology*, 88(5):879, 2003. ISSN 1939-1854. URL https://doi.org/10.1037/0021-9010.88.5.879.

[88] Eva M Pomerantz and Sara G Kempner. Mothers' daily person and process praise: Implications for children's theory of intelligence and motivation. *Developmental Psychology*, 49(11):2040, 2013. ISSN 1939-0599. URL https://doi.org/10.1037/a0031840.

[89] Joseph R. Priester and Richard E. Petty. A research dialogue on mindsets. *Journal of Consumer Psychology*, 26(1): 125–126, 2016. ISSN 1057-7408. URL https://onlinelibrary.wiley.com/doi/abs/10.1016/j.jcps.2015.06.016.

[90] Jimin Pyo and Michael G Maxfield. Cognitive effects of inattentive responding in an mturk sample. *Social Science Quarterly*, 2021. ISSN 0038-4941. URL https://doi.org/10.1111/ssqu.12954.

[91] Dipankar Rai and Chien-Wei Wilson Lin. The influence of implicit self-theories on consumer financial decision making. *Journal of Business Research*, 95:316–325, 2019. ISSN 0148-2963. URL https://doi.org/10.1016/j.jbusres.2018.08.016.

[92] Richard W Robins and Jennifer L Pals. Implicit self-theories in the academic domain: Implications for goal orientation, attributions, affect, and self-esteem change. *Self and identity*, 1(4):313–336, 2002. ISSN 1529-8868. URL https://doi.org/10.1080/15298860290106805.

[93] Jonathan Robinson, Cheskie Rosenzweig, Aaron J Moss, and Leib Litman. Tapped out or barely tapped? recommendations for how to harness the vast and largely unused potential of the mechanical turk participant pool. *PloS one*, 14(12), 2019. URL https://doi.org/10.1371/journal.pone.0226394.

[94] Christine Rzepka and Benedikt Berger. User interaction with ai-enabled systems: a systematic review of is research. In *Proceedings of the International Conference of Information Systems*. ICIS, 2018.

[95] Pericle Salvini, Gaetano Ciaravella, Wonpil Yu, Gabriele Ferri, Alessandro Manzi, Barbara Mazzolai, Cecilia Laschi, Sang-Rok Oh, and Paolo Dario. How safe are service robots in urban environments? bullying a robot. In *19th international symposium in robot and human interactive communication*, pages 1–7. IEEE, 2010. ISBN 1424479908. URL https://doi.org/10.1109/ROMAN.2010.5654677.

[96] Tracy L Sanders, Keith MacArthur, William Volante, Gabriella Hancock, Thomas MacGillivray, William Shugars, and PA Hancock. Trust and prior experience in human-robot interaction. In *Proceedings of the human factors and ergonomics society annual meeting*, volume 61, pages 1809–1813. SAGE Publications Sage CA: Los Angeles, CA, 2017. ISBN 1541-9312. URL https://doi.org/10.1177/1541931213601934.

[97] JL Schleider and HS Schroder. Implicit theories of personality across development: Impacts on coping, resilience, and mental health. *The SAGE Handbook of Personality and Individual Differences. Sage Publications*, 2018. URL http://dx.doi.org/10.4135/9781526470294.n7.

[98] Hans S Schroder, Tim P Moran, M Brent Donnellan, and Jason S Moser. Mindset induction effects on cognitive control: A neurobehavioral investigation. *Biological psychology*, 103:27–37, 2014. ISSN 0301-0511. URL https://doi.org/10.1016/j.biopsycho.2014.08.004.

[99] Karina Schumann and Carol S Dweck. Who accepts responsibility for their transgressions? *Personality and social psychology bulletin*, 40(12):1598–1610, 2014. ISSN 0146-1672. URL https://doi.org/10.1177/0146167214552789.

[100] Yvonne Skipper and Karen Douglas. Is no praise good praise? effects of positive feedback on children's and university students' responses to subsequent failures. *British Journal of Educational Psychology*, 82(2):327–339, 2012. ISSN 0007-0998. URL https://doi.org/10.1111/j.2044-8279.2011.02028.x.

[101] Elizabeth Solberg, Laura EM Traavik, and Sut I Wong. Digital mindsets: Recognizing and leveraging individual beliefs for digital transformation. *California Management Review*, 2020. ISSN 0008-1256. URL https://doi.org/10.1177/0008125620931839.

[102] Birgit Spinath, Frank M. Spinath, Rainer Riemann, and Alois Angleitner. Implicit theories about personality and intelligence and their relationship to actual personality and intelligence. *Personality and Individual Differences*, 35 (4):939–951, 2003. URL https://doi.org/10.1016/S0191-8869(02)00310-0.

[103] Julia G Stapels and Friederike Eyssel. Robocalypse? yes, please! the role of robot autonomy in the development of ambivalent attitudes towards robots. *International Journal of Social Robotics*, pages 1–15, 2021. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-021-00817-2.

[104] Megan K Strait, Cynthia Aguillon, Virginia Contreras, and Noemi Garcia. The public's perception of humanlike robots: Online social commentary reflects an appearance-based uncanny valley, a general fear of a "technology takeover", and the unabashed sexualization of female-gendered robots. In *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, pages 1418–1423. IEEE, 2017. ISBN 1538635186. URL https://doi.org/10.1109/ROMAN.2017.8172490.

[105] Maya Tamir, Oliver P John, Sanjay Srivastava, and James J Gross. Implicit theories of emotion: Affective and social outcomes across a major life transition. *Journal of personality and social psychology*, 92(4):731, 2007. ISSN 1939-1315. URL https://doi.org/10.1037/0022-3514.92.4.731.

[106] Xiang Zhi Tan, Marynel Vázquez, Elizabeth J Carter, Cecilia G Morales, and Aaron Steinfeld. Inducing bystander interventions during robot abuse with social mechanisms. In *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, pages 169–177, 2018. URL https://doi.org/10.1145/3171221.3171247.

[107] Tuuli Turja and Atte Oksanen. Robot acceptance at work: A multilevel analysis based on 27 eu countries. *International Journal of Social Robotics*, 11(4):679–689, 2019. ISSN 1875-4805. URL https://doi.org/10.1007/s12369-019-00526-x.

[108] Sheryl L Walter, Scott E Seibert, Daniel Goering, and Ernest H O'Boyle. A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, 34(4):425–452, 2019. ISSN 1573-353X. URL https://doi.org/10.1007/s10869-018-9552-y.

[109] Ricarda Wullenkord and Friederike Eyssel. The influence of robot number on robot group perception—a call for action. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(4):1–14, 2020. ISSN 2573-9522. URL https://doi.org/10.1145/3394899.

[110] David S Yeager and Carol S Dweck. What can be learned from growth mindset controversies? *American Psychologist*, 75(9):1269, 2020. ISSN 1935-990X. URL https://doi.org/10.1037/amp0000794.

[111] David S Yeager, Adriana S Miu, Joseph Powers, and Carol S Dweck. Implicit theories of personality and attributions of hostile intent: A meta-analysis, an experiment, and a longitudinal intervention. *Child development*, 84(5):1651–1667, 2013. ISSN 0009-3920. URL https://doi.org/10.1111/cdev.12062.

[112] Kumar Yogeeswaran, Jakub Złotowski, Megan Livingstone, Christoph Bartneck, Hidenobu Sumioka, and Hiroshi Ishiguro. The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. *Journal of Human-Robot Interaction*, 5(2):29–47, 2016. ISSN 2163-0364. URL https://doi.org/10.5898/JHRI.5.2.Yogeeswaran.

[113] Eric A Yorkston, Joseph C Nunes, and Shashi Matta. The malleable brand: The role of implicit theories in evaluating brand extensions. *Journal of Marketing*, 74(1):80–93, 2010. ISSN 0022-2429. URL https://doi.org/10.1509/jmkg.74.1.80.

[114] Elizabeth S Zack, John Kennedy, and J Scott Long. Can nonprobability samples be used for social science research? a cautionary tale. In *Survey Research Methods*, volume 13, pages 215–227, 2019. ISBN 1864-3361. URL https://doi.org/10.18148/SRM/2019.V13I2.7262.

[115] Jakub Złotowski, Kumar Yogeeswaran, and Christoph Bartneck. Can we control it? autonomous robots threaten human identity, uniqueness, safety, and resources. *International Journal of Human-Computer Studies*, 100:48–54, 2017. ISSN 1071-5819. URL https://doi.org/10.1016/j.ijhcs.2016.12.008.