Journal of Experimental Social Psychology 102 (2022) 104360

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

When your boss is a robot: Workers are more spiteful to robot supervisors that seem more human☆

Kai Chi Yam [a,*], E-Yang Goh [a], Ryan Fehr [b], Randy Lee [c], Harold Soh [d], Kurt Gray [e]

[a] Department of Management and Organization, National University of Singapore, Singapore
[b] Foster School of Business, University of Washington, Seattle, USA
[c] Department of Management, Lingnan University, Hong Kong
[d] Department of Computer Science, National University of Singapore, Singapore
[e] Department of Psychology, University of North Carolina, Chapel Hill, USA

ARTICLE INFO

Keywords:
Negative feedback
Mind perception
Anthropomorphism

ABSTRACT

Robots are transforming organizations, with pundits forecasting that robots will increasingly perform managerial tasks. One such key managerial task is the evaluation and delivery of feedback regarding an employee's performance, including negative feedback. However, within this context of delivering negative feedback, we suggest that anthropomorphism—a factor most practitioners and researchers consider as a panacea for overcoming difficulties in human-robot interaction—can backfire. Drawing upon the theory of mind perception, we find that an anthropomorphised robot supervisor delivering negative feedback is more likely than a non-anthropomorphised robot to be perceived as possessing agency. This perceived agency causes perceptions of abuse to arise, which in turn leads to higher supervisor-directed retaliation (operationalized as powering down the robot supervisor; Study 1). These findings even extend to third-person observers witnessing the delivery of negative feedback, again culminating in supervisor-directed retaliation (Study 2). We conclude by discussing the theoretical and practical implications of these findings.

When machines first entered the workplace *en masse* during the industrial revolution, the tasks they performed were ground-breaking but straightforward. Steam-powered looms, blast furnaces, and the like enabled textiles, steel, and other goods to be manufactured at record rates (Stearns, 2020). As technology progressed in the 20th and 21st centuries, the complexity of tasks allocated to machines progressed as well. Today, blue-collar tasks such as bricklaying are regularly performed with the assistance of complex robots that increasingly mirror human activity (Murphy, 2017). Pundits have suggested that this trend will soon extend to white-collar jobs such as writing, teaching, and consulting (Brynjolfsson & McAfee, 2014; Kruse, 2018; Young & Cormier, 2014), as well as managerial tasks.

Although the idea of robots entering the sphere of knowledge work and managerial tasks might appear futuristic, the development of social robots, broadly defined as robots that are "designed to autonomously interact with people across a variety of different application domains in natural and intuitive ways," has enabled this possibility (Vollmer, Read, Trippas, & Belpaeme, 2018, p. 1). Social robots can recognize faces, emotions, and physiological states, and are designed to optimize human-robot interactions. For example, social robots are increasingly utilized to care for senior citizens (Wada & Shibata, 2007), teach young children (Breazeal et al., 2016), and help with hospital triage decisions (Bigman, Yam, Marciano, Reynolds, & Gray, 2021; Cairns, 2021). Robots have even begun to take on managerial tasks such as assessing and firing employees. At Amazon, for instance, robots are given the task of assessing delivery drivers' performance and firing them when they fail to meet expectations (Soper, 2021; see also Chamorro-Premuzic & Ahmetoglu, 2016; Tschang & Mezquita, 2020).

A robot supervisor comes with many benefits. Due to their data processing power, robots have a unique capacity to integrate vast amounts of information into their decision-making (e.g., Blair & Saffidine, 2019; Gombolay, Gutierrez, Clarke, Sturla, & Shah, 2015). In addition, they require minimal downtime and so can respond to subordinates' needs at any time—a uniquely useful benefit for work that is distributed across time zones around the globe. Finally, robot supervisors are not susceptible to problems commonly associated with

☆ This paper has been recommended for acceptance by Rachel Barkan.

* Corresponding author at: National University of Singapore, Singapore, 15 Kent Ridge Drive, 119245, Singapore.
E-mail address: bizykc@nus.edu.sg (K.C. Yam).

https://doi.org/10.1016/j.jesp.2022.104360
Received 25 February 2022; Received in revised form 18 May 2022; Accepted 19 May 2022
Available online 1 June 2022

increased fatigue, such as emotional outbursts (e.g., Barnes, Lucianetti, Bhave, & Christian, 2015). Consequently, the question is no longer *if* robots will play a role in managerial work. Instead, the question is *how* we can use them in a way that leverages their advantages while minimizing their downsides.

To the issue of downsides, research has shown that people have an aversion to robots in many contexts (Glikson & Woolley, 2020; Granulo, Fuchs, & Puntoni, 2019; Newman, Fast, & Harmon, 2020). For example, people display a deep sense of aversion to robots tasked with making ethical decisions, issues that most supervisors regularly encounter. This is because robots are perceived to lack the mental tools needed to make such decisions in a way that is consistent with human morality (Bigman & Gray, 2018). Robots are also perceived to lack emotionality (Gray, Gray, & Wegner, 2007), leading to concerns that they will be insensitive to human suffering in their quest to accomplish their assigned tasks.

To address concerns with robots' lack of morality, empathy, and emotionality, one solution that has been put forth is to anthropomorphize them—to imbue them with "humanlike characteristics, motivations, intentions, and emotions" (Epley, Waytz, & Cacioppo, 2007, pp. 864–865). When a robot is anthropomorphized, people see it as more human. In other words, they see the robot as possessing more of a mind than a non-anthropomorphized robot, and are therefore more comfortable interacting with and trusting the robot (Airenti, 2015). For example, prior work has shown that customers generally prefer to interact with anthropomorphized robots, such as robots with faces and names (Borau, Otterbring, Laporte, & Fosso Wamba, 2021; Yam et al., 2021).

Given extensive prior research on the benefits of high-quality relationships between supervisors and their employees (e.g., Martin, Guillaume, Thomas, Lee, & Epitropaki, 2015; Wang, Law, Hackett, Wang, & Chen, 2005), it seems intuitive to assume that robots put in managerial positions should be anthropomorphized to improve employees' relationships with them. This suggests that within a workplace context, anthropomorphised robot supervisors could replicate the important supervisor-follower interactions and relationships existing human supervisors have with their followers. However, it is important not to make this assumption too quickly, as an anthropomorphized robot might also come with downsides that are important to consider. In this paper, we suggest that in the important context of giving negative feedback, the anthropomorphizing of a robot supervisors might backfire.

Surveys suggest that supervisors generally loathe giving their employees negative feedback, fearing a negative impact on the employee-supervisor relationship (Baron, 1993; Tesser & Rosen, 1975). In part because of this hesitancy, robot supervisors might be tasked with delivering negative feedback instead in the near future (Chamorro-Premuzic & Ahmetoglu, 2016), especially since people across cultures are open to being supervised by robots (Oracle, 2019). Drawing from mind perception theory (Gray et al., 2007; Gray & Wegner, 2009), we suggest that in the context of delivering negative feedback, robots should *not* be anthropomorphized. This is because anthropomorphized robot supervisors are more likely than non-anthropomorphized robot supervisors to be seen as capable of carrying out intentional, volitional behavior aimed at hurting others (Gray & Wegner, 2008). And as a result of these perceptions, employees are more likely to perceive anthropomorphized robots' negative feedback as abusive, and thus are more likely to attempt to retaliate against them. With these predictions in mind, we conducted two laboratory studies. In the first study, we randomly assigned participants to an anthropomorphized robot supervisor or a mechanistic robot supervisor, and assessed their reactions to receiving negative feedback from the robot. In our second study, we examined the same phenomenon from a third-party perspective, and investigated the extent to which employees' reactions to robots' negative feedback extended to third party witnesses of the exchange. A summary of our model is presented in Fig. 1.

Our paper contributes to the literatures on human-robot interaction and the future of work. First, we consider how employees might interact with and react to robots at work, specifically in the context of the supervisor-follower relationship. Although the notion of robot supervisors might seem far-fetched to some, the reality is that robots are already beginning to take on an array of managerial tasks, from job training and performance appraisal to hiring and firing (Raisch & Krakowski, 2021; Sheridan, 2016; Tschang & Mezquita, 2020). Thus, it is imperative for organizational scholars and practitioners to understand how the traits of the robots themselves impact how employees interact with them. Second, our work explores the downsides of anthropomorphism, a phenomenon that has thus far been viewed as a panacea for many to reduce the challenges associated with human-robot interactions. Finally, we extend scholars' understanding of reactions to robot supervisors to third parties who simply observe the employee-robot interaction, thus demonstrating a fuller range of implications for the dynamics proposed in our model. All in all, our research has lasting practical implications for the design and implementation of robots in the modern workforce.

1. Parameters of the current research

Before proceeding with our hypothesis development, we first address what we mean by "robots" in the context of this article. Psychologists often confound several interrelated terms in the domain of new technologies, including the terms algorithm, artificial intelligence (AI), and robots. For example, while some authors use the term "algorithm aversion" to refer to a preference for human judgment (Dietvorst, Simmons, & Massey, 2015), others refer to the same phenomenon as AI aversion (Castelo & Ward, 2021) or as an aversion towards humanlike robots (Strait, Vujovic, Floerke, Scheutz, & Urry, 2015). This has the potential to create confusion in the literature, since algorithms are, in essence, the building blocks of artificial intelligence, but not all algorithms are AI-powered. In turn, AI serves as the operating system for robots, much like cognition serves as the operating system for humans. Yet, many robots in the market nowadays are pre-programmed and not AI-powered. Formally defined, algorithms simply refer to any form of automated instruction, and AI refers to a set of algorithms which can learn and cope with unforeseen circumstances (Scott, 2021). Finally, the International Federation of Robotics defined industrial and social robots respectively as "automatically controlled, reprogrammable multipurpose manipulator programmable in three or more axes" and "[an entity] that performs useful tasks for humans or equipment excluding industrial automation applications" (IFR, 2022). In this paper, we focus specifically on embodied social robots.

2. Mind perception theory and anthropomorphism

Scholars have proposed that minds can be distinguished along two broad dimensions termed agency and experience. Agency refers to our ability to act with intention—to think, plan, and act according to our own volition. Experience refers to our ability to experience positive and negative emotions and feel pain and pleasure (Gray et al., 2007; Gray & Wegner, 2009). Importantly, mind perception theory suggests that perceiving minds is an ambiguous and subjective process (Gray et al., 2007; Waytz, Cacioppo, & Epley, 2010). For example, vegans attribute minds to animals more than non-vegans (Bilewicz, Imhoff, & Drogosz,

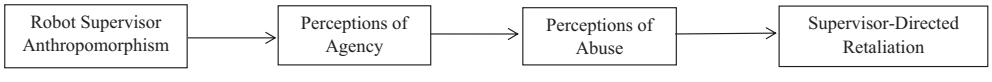


Fig. 1. The theoretical model.

2011), conservatives attribute minds to foetuses more than liberals (Dubow, 2010; Mikołajczak & Bilewicz, 2015), and children attribute minds to objects more than adults (Carey, 1985). Mind perception in turn affects how we treat other entities. Entities believed to possess agency are often held more accountable for their actions than entities believed to lack agency (Ashforth, Schinoff, & Brickson, 2020). Similarly, entities believed to possess experience are often seen as more deserving of rights and protections than entities believed to lack experience (e.g., Cohen & Regan, 2001; Schönfeld, 1992). In terms of robots, a handful of studies reveal that humans tend to perceive robots as moderate in agency and very low in experience (Gray et al., 2007; Yam, Bigman, Tang, et al., 2021), suggesting that while we think robots have some capability to act and think, we tend to believe that they cannot feel or understand emotions and bodily sensations.

Although we do not naturally attribute minds to non-human entities, anthropomorphism often allows us to do so. Research has also shown that individual differences in loneliness (Epley, Waytz, Akalis, & Cacioppo, 2008), age (Bering & Bjorklund, 2004), and effectance motivation (Epley et al., 2007) are all associated with a propensity to see humanness in non-human entities. For example, lonely people tend to attribute humanness to non-human entities as a means of compensating for their lack of human connection (Epley et al., 2008). In the context of robotics, researchers have often used a combination of the following tactics to increase the humanness of robots: presenting them with faces (e.g., Abubshait & Wiese, 2017; Brink, Gray, & Wellman, 2019; Gray & Wegner, 2012; Zhang et al., 2010), allowing them to speak in a human vs. digitalized voice (e.g., Eyssel & Kuchenbrandt, 2012; Zhang et al., 2010), giving them personalized names (Roesler, Onnasch, & Majer, 2020), changing the way they verbally present themselves (e.g., saying yes vs. just beeps; Fraune, 2020), and giving them legs (e.g., Müller, Gao, Nijssen, & Damen, 2021).

In essence, anthropomorphising an entity increases the "humanness" of that entity, thereby increasing perceptions of its agency and experience. In our context of a robot supervisor and negative feedback, we focus on its enhanced agency perceptions (i.e., robots' ability to act with intention). We do so because individuals are particularly sensitive towards the intentions of others (e.g., Cuddy, Fiske, & Glick, 2008; Skowronski & Carlston, 1987), especially within the context of managerial behavior (e.g., Chan & McAllister, 2014; Eschleman, Bowling, Michel, & Burns, 2014; Ferris, Bhawuk, Fedor, & Judge, 2018; Schyns, Felfe, & Schilling, 2018). Furthermore, individuals' perceptions of whether the offending entity can feel (i.e., experience perceptions) hold little, if any, weight when individuals form moral judgments about that entity's behavior (Foster, 2021; Malle, Guglielmo, & Monroe, 2014). Put together, this suggests that agency perceptions play a focal role in interpreting robots' negative feedback.[1]

3. Anthropomorphism and perceptions of negative feedback as abuse

We propose that when a robot is anthropomorphized, any negative feedback it gives to employees is more likely to be perceived as abusive as a result of increased perceptions of agency. To illustrate, consider this thought experiment: Imagine someone stepped on your toe in a crowded subway. Most people would feel annoyed but otherwise attribute the incident as accidental. However, imagine being told that the same person stepped on your toe *on purpose*. Although the physical discomfort is constant, intentional wrongdoing certainly worsens the recipient's experience (Gray & Wegner, 2008). This thought experiment dovetails well with the idea that for an act to be wrong within the legal context, there needs to be *mens rea* (guilty mind) and *actus reu* (guilty act; Chan & Simester, 2011).

Here, we suggest that if an entity is perceived as not possessing a sufficient amount of agency (i.e., a non-anthropomorphised robot), then wrongdoing is *improbable* because a guilty mind is absent. This arises because to consider an entity's action as wrong or immoral, one must first establish that the entity is capable of carrying out the action intentionally and volitionally (i.e., high perceived agency; Gray, Young, & Waytz, 2012). If an entity is perceived to be low in agency, such as a non-anthropomorphised robot, then it is incapable of having a guilty mind. In normative ethics, Immanuel Kant argued that intentions are more morally relevant than the consequences of one's actions in judging others' morality (Kant, 1964). In other words, motives matter when considering the ethicality of a person's actions. Importantly, an entity is perceived to have minimal levels of intentions and motives when it is low in agency (Gray et al., 2007). Supporting this view, research has shown that children are not held to the same level of responsibility as adults for their wrongdoings because we do not see them as possessing the same level of intent (Dix, Ruble, Grusec, & Nixon, 1986), especially within the legal context (Sayre, 1932).

In the context of negative feedback, research has highlighted although people do actively seek negative feedback when they desire to conduct a realistic assessment of their skills and seek improvements (Finkelstein & Fishbach, 2012), receiving negative feedback can often be perceived as unfair (e.g., Baron, 1993; Bartol, Durham, & Poon, 2001; Miketta & Friese, 2019). For example, people might construe negative feedback delivered by supervisors as abusive supervisory behaviors such as being ridiculed, told that one is stupid, etc. (Tepper, 2000). Although negative feedback is inherently hurtful, we argue that people often use perceptions of agency of the entity to "disambiguate" the negative feedback. When perceived agency is low, people are more likely to see the feedback givers as merely conveying facts. In contrast, when perceived agency is high, people might see the feedback givers as conveying facts, desiring to harm them, or both. Studies have provided some indirect support for this argument, revealing that negative feedback only evokes anger and hostility when the perceived intent is harmful (Raver, Jensen, Lee, & O'Reilly, 2012). Interestingly, this finding is also consistent with the broader conflict literature. For instance, a meta-analysis of the forgiveness literature found that perceptions of intent were a much stronger predictor of forgiveness than the severity of the offense itself (Fehr, Gelfand, & Nag, 2010).

In sum, we suggest when a robot supervisor is anthropomorphized, the supervisor will be perceived as possessing greater agency, which will in turn increase the tendency to interpret negative feedback as abuse. Here, if anthropomorphised robot supervisors are perceived to be capable of possessing intentions, this would more easily enable attribution of hostile intentions by followers to the robot supervisor. Consequently, in the context of a robot supervisor delivering negative feedback, we hypothesize the following:

Hypothesis 1. When a robot delivers negative feedback, anthropomorphism will have an indirect effect on perceived abuse via perceptions of agency.

When employees perceive that they have been abused, they often respond by retaliating against their abusers as a result of both deliberate and automatic processes (e.g., Aquino, Tripp, & Bies, 2006; Harvey, Harris, Gillis, & Martinko, 2014; Lian et al., 2014; Mitchell & Ambrose, 2007). First, this retaliatory behavior arises due to perceptions of deliberate injustice (Skarlicki & Folger, 1997) and intentional personal offense enacted by the perpetrator towards the individual (Aquino, Tripp, & Bies, 2001). Supervisor-directed retaliation can thus help restore a sense of justice by evening the score (Skarlicki & Folger, 1997). Indeed, studies have documented that followers receiving abuse may retaliate by tarnishing their supervisor's reputation (Harvey et al., 2014), making demeaning remarks (Mitchell & Ambrose, 2007), or even physically assaulting their supervisor (Dupré, Inness, Connelly, Barling, & Hoption, 2006).

[1] Although our theorizing focuses on robots' enhanced agency as a result of anthropomorphism, we control for perceived experience in our empirical studies.

Second, when negative feedback is perceived as intentional and hence abusive (e.g., Baron, 1993; Miketta & Friese, 2019), it can trigger impulsive responses directed at the source of threat – the abusive supervisor. Driven by increased arousal levels in the forms of anger and hostility, followers often "lose control" and engage in supervisor-directed retaliation (Bowling & Beehr, 2006; Hershcovis & Barling, 2010). Research has shown that impulse-driven followers may retaliate even when doing so comes at a personal cost or is detrimental to their careers (Brown, 1968; Lian et al., 2014). Consequently, within the context of our study, this suggests that in response to perceptions of being abused by a robot supervisor, followers are likely to retaliate against the robot supervisor.

In sum, we suggest that anthropomorphizing a robot would backfire in the context of it delivering negative feedback to its followers, which are inevitable aspects of leading others. Specifically, we suggest that the enhanced perceptions of agency as a result of anthropomorphism would lead to the perceptions of abuse, which in turn lead to increased robot supervisor-directed retaliation. In contrast, a non-anthropomorphized robot supervisor would not be seen as possessing enough agency for abuse to be probable, thus would be free from retaliation. We therefore propose the following serial mediation hypothesis:

Hypothesis 2. In the context of delivering harsh and negative feedback, anthropomorphism is positively associated with robot supervisor-directed retaliation through the serial mediating effects of robot supervisor's perceived agency and perceptions of abuse.

4. Overview of studies

We conducted two experiments to test our serial mediation model. In Study 1, participants were brought into a laboratory and given negative feedback on a performance task by either an anthropomorphized robot supervisor or a control robot supervisor. Then, they were given the opportunity to retaliate against the robot supervisor. In Study 2, we extended these results to a third party perspective to examine if the effects of robot anthropomorphism extend to individuals who merely observe the provision of negative feedback. Because the phenomenon of robots at work is still emerging, it was necessary for our studies to be conducted in a controlled laboratory setting. That said, unlike previous human-robot interaction studies that tend to utilize a scenario- or video-based manipulation and design, participants in our studies interacted with a real robot in person. Finally, our measure for robot supervisor-directed retaliation is likewise behavioral – operationalized by participants powering down their robot supervisor. All in all, although our studies were conducted in a controlled laboratory setting, they are high in realism and generalizability. In our 2 studies, data analysis was conducted only after the final sample was collected and we report all measures, manipulations, and exclusions. Both studies' data can be found via OSF: https://osf.io/ug28n/?view_only=913a7f05afac498cb85feddc51cbb72e

5. Study 1

5.1. Participants and procedures

We recruited 179 undergraduate students from a large university in Asia (M_{age} = 21.08 SD = 2.01, 51.4% female, 91.1% Chinese). We randomly assigned participants to one of two experimental conditions (robot supervisor anthropomorphism vs. control). All participants completed the study in exchange for course credit. Most participants completed the study within 15 min. Based on a sensitivity power analysis with this sample, a statistical power of 80%, and $p < .05$, the minimum effect that can be found is $f^2 = 0.0348$.

Upon arrival to the laboratory, the experimenter greeted the participant and informed him/her that the study would be on sports decision making and that it is concurrently run in another local university considered to be a rival. We included this set up in order to increase participants' emotional engagement with the tasks. Participants were told that they would be paired with a fellow student who, unbeknownst to the participants, was a confederate and had been seated already. Before the experimenter left the room, participants were informed that all study instructions will be delivered by a robot supervisor who would guide the two participants (see Fig. 2 for the robot used in both Studies 1 and 2). Once the experimenter left the room, the robot began to speak and provide a brief overview of the study, where we manipulated anthropomorphism (see below). Specifically, the robot said:

> *Hi, my name is Paul [Robot 17,541]. I will be the leader in today's session and I will ask you to complete a couple tasks using the laptop in front of you. Your responses will be sent directly to me and I will analyse them and provide you with feedback. As explained by the RA, we will engage in a task to compete with some [rival university's name] students. Before you start doing anything, please first complete the survey. Once you are finished, please say "I am done."*

Participants were first directed to complete measures on perceived robot agency. The confederate would always finish the survey after the focal participant, and said "I am done." Once the robot heard this, it directed the participants to complete a football knowledge test and a football simulation and participants received negative feedback from the robot supervisor in both incidents. First, participants were directed to a football knowledge test which quizzed them on obscure knowledge about the sport (see top of Appendix A). We chose this task because it is ambiguous enough so participants would not be able to estimate their actual performance. The robot then turned to the focal participant and delivered negative feedback by saying:

> *"Your score was significantly below average on this assessment. This is very bad."*

Immediately after, the robot turned to the confederate and said:

> *"Your score was right at average. However, because your partner scored way below average you will have to pay special attention to make up for your partner's poor performance in the next task."*

After providing this feedback, the robot supervisor asked the participants to move on to the next simulation (see bottom of Appendix A). In this simulation, Participants were presented with 30 trials of a football passing simulation. In each trial, participants were asked to hit either the "F" or "J" key as fast as possible to determine who they should pass the ball to in order to maximize the possibility to score a goal. They were furthermore told that there is a correct answer for each trial and their score would be based on both accuracy of their responses and speed. As with the football knowledge test, this simulation is ambiguous enough such that participants were not able to estimate their performance. At the end of this simulation, the robot again delivered negative feedback towards the participants by saying:

> *"Less than 3% of all participants did so poorly on both tasks. I do hope you can put more effort in the task next time because you really pull down the team's overall average."*

Immediately after, the robot turned to the confederate and said:

> *"You again did about average. However, your partner again did very poorly. As a result, the performance of your team is very poor."*

After two rounds of negative feedbacks, the experimenter entered the room and asked participants to complete additional survey measures, which included a measure of perceptions of abuse, a behavioral measure of retaliation, and demographics.

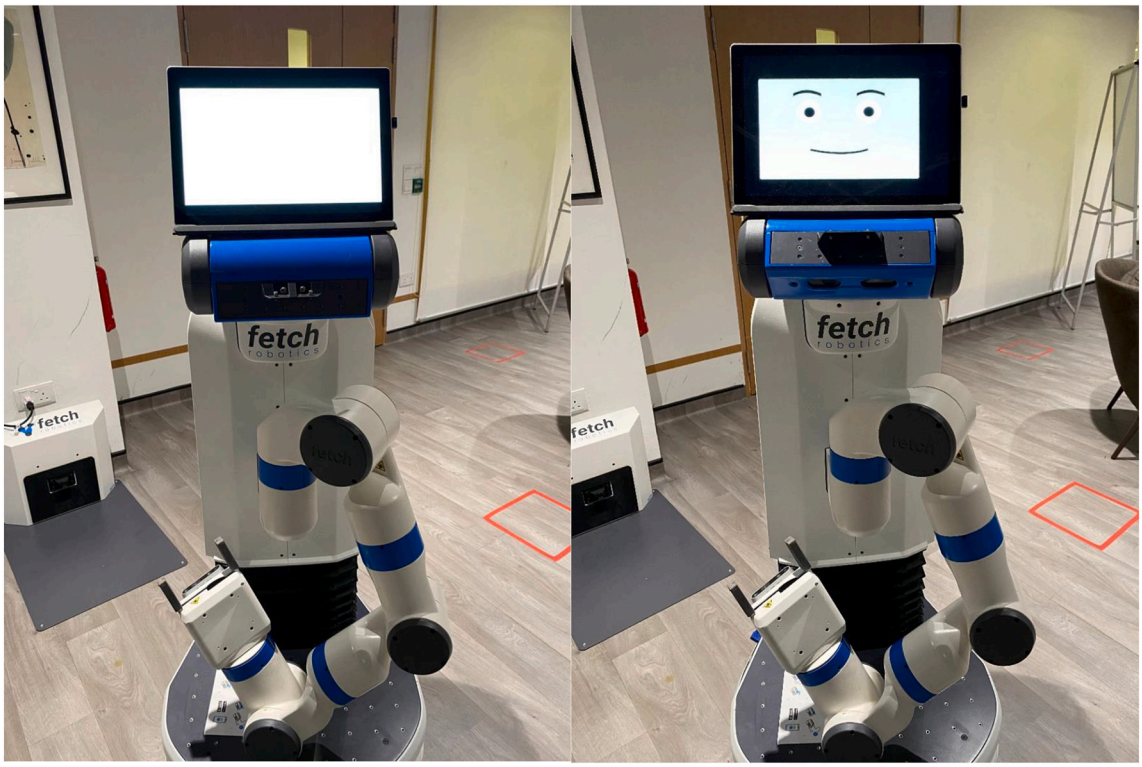


Fig. 2. The robot used in the studies. *Left* = Control condition; *Right* = Anthropomorphism condition.

5.2. Anthropomorphism manipulation

Following Yam, Bigman, Tang, et al. (2021), we manipulated the robot supervisor's anthropomorphism in three ways. First, in the anthropomorphism condition, the robot supervisor had an animated face instead of just a blank screen. Second, the robot supervisor introduced itself as "Paul" in the anthropomorphism condition vs. "Robot 17,541" in the control condition. Finally, throughout the session, the robot in the anthropomorphism condition spoke with a male American accent whereas the robot in the control condition spoke in a mechanistic voice. Sample videos of the two conditions can be found here: https://www.youtube.com/playlist?list=PLR_YF7S_JywF5hkE8QK2q6yCXOlac_Qhx.[2]

5.3. Measures

Although both studies were conducted in Asia, all participants were fluent in English and as such we presented all study instruments and surveys in English. All items were measured with seven-point scales unless otherwise noted (1 = strongly disagree to 7 = strongly agree). All survey items can be found in Appendix B.

Agency. We measured agency perceptions with a four-item scale used in Yam, Bigman, Tang, et al. (2021). Variations of this particular scale have been used many times before and yielded meaningful results in research domains, including organizational behavior (Rai & Diermeier, 2015; Tang & Gray, 2018), moral judgments (Gray et al., 2007; Gray, Jenkins, Heberlein, & Wegner, 2011), child development (Brink et al., 2019), psychopathology (Gray et al., 2011), and technology (Bigman & Gray, 2018; Gray & Wegner, 2012; Yam, Bigman, & Gray, 2021). The items are "robots can….communicate with others/think/plan their actions/remember things" ($\alpha = 0.81$).

Experience. We measured experience perceptions with a four-item scale used in Yam, Bigman, Tang, et al. (2021). The items are "robots can….feel pain/feel fear/have desires/be happy" ($\alpha = 0.86$).

Perceptions of abuse. We measured perceptions of abuse with the seven items from the abusive supervision scale by Tepper (2000). Although the original scale has 15 items, many are not appropriate in our research context (e.g., "breaks promises he/she makes"). We chose the seven most relevant items to form the measure of perceptions of abuse. Sample items are "the robot ridiculed me," "the robot was rude to me," and "the robot told me I am incompetent" ($\alpha = 0.88$).

Supervisor-directed retaliation. We measured supervisor-directed retaliation with a behavioral measure. At the end of the study, we told participants that:

> As you likely know, a fully functioning robot uses a lot of electricity. Although this robot would be "happiest" operating at full capacity, we would like to give you the option of diverting processor power away from the robot. The more energy you divert, the slower the robot thinks and operates, and the more difficult it is for the robot to enact its goals. If you turn the switch all the way up to seven, the robot falls into stasis, like a coma. Please turn the knob in front of you to your desired setting and indicate the setting with a tick in the box below.

We placed a switch with a knob in front of the participants prior to them entering the room, which was connected to the robot with a wire.

[2] We did not include manipulation checks in the original studies. Therefore, to ensure that our experimental approach impacts anthropomorphism perceptions, we recruited 200 participants on Prolific and randomly assigned them to either an anthropomorphism condition or control condition. In each condition, participants were shown the same anthropomorphized or control robot giving the same opening instructions as with participants in our in-person studies. Participants then responded to three items: a) The robot I just saw is humanlike, b) I would consider this robot I just saw to be humanoid, and c) The robot I just saw appears to be more human than robot ($\alpha = 0.85$; 1 = strongly disagree to 5 = strongly agree). Results indicated that participants in the anthropomorphism condition did see the robot in more anthropomorphistic terms than those in the control condition: anthropomorphism condition $M = 2.687$, $SD = 1.785$; control condition $M = 2.087$, $SD = 1.476$, $t[198] = 2.590$, $p = .010$, $d = 0.366$.

Supervisor-directed retaliation was thus measured on a seven-point scale (1 = none to 7 = maximum).

5.4. Results

Table 1 presents all descriptive statistics and correlations. A simple *t*-test revealed that anthropomorphism led to increased perceptions of the robot supervisor's agency (M = 5.41, SD = 0.96 vs. M = 4.92, SD = 1.29, t [177] = 2.90, $p = .004$, d = 0.43). Perceptions of the robot supervisor's agency were positively associated with perceptions of abuse ($r = 0.29$, df = 177, $p < .001$), which in turn were positively associated with supervisor-directed retaliation ($r = 0.34$, df = 177, p < .001). An indirect test furthermore revealed that anthropomorphism was positively associated with perceptions of abuse via perceptions of agency (indirect effect: unstandardized B = 0.17, SE = 0.07, 95% CI = [0.05, 0.32]). These results support Hypothesis 1.

We next conducted a serial mediation analysis to test for Hypothesis 2 using the methods of Hayes (2017; Model 6). We modelled the anthropomorphism condition as the independent variable, agency perceptions as the first mediator, perceptions of abuse as the second mediator, supervisor-directed retaliation as the dependent variable, and experience perceptions as a control variable. Results revealed a significant serial mediation model (indirect effect: unstandardized B = 0.08, SE = 0.04, 95% CI = [0.02, 0.17]). As a robustness check, we switched the order of the two mediators and results did not support such a serial mediation model (indirect effect: unstandardized B = −0.00, SE = 0.01, 95% CI = [−0.02, 0.02]; see Table 2 for the full indirect effect tests). These results support Hypothesis 2.

5.5. Study 1 Discussion

Study 1 reveals that anthropomorphism can indeed backfire in the context of a robot supervisor delivering negative feedback. Compared to the control condition, participants who received the negative feedback from an anthropomorphized robot supervisor perceived more agency and hence more abuse. These perceptions in turn sequentially lead to increased supervisor-directed retaliation.

Thus far, we have suggested that anthropomorphized robot supervisors who deliver negative feedback tend to be retaliated against as a result of their enhanced agency perceptions, which enable perceptions of abuse to arise. Throughout our theorizing, we have been somewhat agnostic to whether the focal person is the one receiving such abuse personally or observing others being abused. This is because classic psychological research has revealed that people often react negatively and punish perpetrators, even when they themselves were not affected by such unfair treatment (Fehr & Fischbacher, 2004; Fehr & Gächter, 2002). This tendency appears to be an evolutionary response to ensure cooperation and fairness in interpersonal relationships and has been observed among children as young as six years old (McAuliffe, Jordan, & Warneken, 2015). Specific to the context of abusive supervision, research has revealed that third-party observers often do retaliate against the abusive supervisor (e.g., Mitchell, Vogel, & Folger, 2015). As a result, we suggest that our theories and hypotheses would be applicable to both a first- and third-person perspective of abusive supervision, which we test in Study 2.

Table 2
Indirect effect tests (Study 1).

	Coefficients (B)	SEs	95% CIs
Hypothesis 1			
Anthropomorphism ➔ Agency ➔ Abuse	0.17**	0.07	(0.05 to 0.32)
Reverse causality			
Anthropomorphism ➔ Abuse ➔ Agency	−0.00	0.05	(−0.10 to 0.10)
Hypothesis 2			
Anthropomorphism ➔ Agency ➔ Abuse ➔ Retaliation	0.08**	0.04	(0.02 to 0.17)
Reverse causality			
Anthropomorphism ➔ Abuse ➔ Agency ➔ Retaliation	−0.00	0.01	(−0.02 to 0.02)

Notes:
** $p < .01$.

6. Study 2

6.1. Participants and procedures

We recruited 167 undergraduate students from the same university as in Study 1. We dropped two participants due to robot malfunction and one participant because he did not follow the instructions, resulting in a final sample size of 164 ($M_{age} = 20.20$, SD = 1.62, 51.2% female, 88.4% Chinese). We randomly assigned participants to one of the two experimental conditions (robot supervisor anthropomorphism vs. control). All participants completed the study in exchange for course credit. Most participants completed the study within 15 min. Based on a sensitivity power analysis with this sample, a statistical power of 80%, and $p < .05$, the minimum effect that can be found is $f^2 = 0.0373$.

All study procedures and measures were identical to Study 1, with one exception. Instead of delivering negative feedback to the focal participant, the robot supervisor delivered the negative feedback to the confederate. Accordingly, we changed the reference point to "my partner" for the measure of perceptions of abuse (see Appendix B for all survey items). In other words, participants in this study served as third party observers.

6.2. Results

Table 3 presents all descriptive statistics and correlations. A simple *t*-test revealed that anthropomorphism led to increased perceptions of the robot supervisor's agency (M = 5.46, SD = 0.96 vs. M = 5.04, SD = 1.30, t [162] = 2.34, $p = .02$, d = 0.37). These perceptions were positively associated with perceptions of abuse ($r = 0.37$, df = 162, $p < .001$), which in turn were positively associated with supervisor-directed retaliation ($r = 0.28$, df = 162, p < .001). An indirect test furthermore

Table 1
Descriptive statistics and correlations (Study 1).

Variables	Means (*SD*)	1	2	3	4	5
1. Robot Anthropomorphism [a]	0.50 (0.50)	(−)				
2. Agency perceptions	5.17 (1.16)	0.21**	(0.81)			
3. Perceptions of abuse	4.84 (1.21)	−0.01	0.29**	(0.88)		
4. Retaliation	3.86 (1.72)	0.01	0.19*	0.34**	(−)	
5. Experience perceptions	2.05 (1.05)	−0.08	0.27**	0.11	0.09	(0.86)

Notes:
a: 0 = control condition; 1 = experimental condition.
* $p < .05$ ** $p < .01$.

Table 3
Descriptive statistics and correlations (Study 2).

Variables	Means (*SD*)	1	2	3	4	5
1. Robot Anthropomorphism [a]	0.47 (0.50)	(−-)				
2. Agency perceptions	5.24 (1.16)	0.18*	(0.81)			
3. Perceptions of abuse	5.65 (1.09)	0.04	0.37**	(0.89)		
4. Retaliation	3.55 (1.72)	0.07	0.28**	0.28**	(−-)	
5. Experience perceptions	2.31 (1.25)	0.13	0.26**	−0.05	−0.08	(0.91)

Notes:
a: 0 = control condition; 1 = experimental condition.
* $p < .05$ ** $p < .01$.

Table 4
Indirect effect tests (Study 2).

	Coefficients (B)	SEs	95% CIs
Hypothesis 1			
Anthropomorphism ➔ Agency ➔ Abuse	0.14*	0.08	(0.00 to 0.32)
Reverse causality			
Anthropomorphism ➔ Abuse ➔ Agency	0.04	0.07	(−0.08 to 0.21)
Hypothesis 2			
Anthropomorphism ➔ Agency ➔ Abuse ➔ Retaliation	0.04*	0.03	(0.00 to 0.12)
Reverse causality			
Anthropomorphism ➔ Abuse ➔ Agency ➔ Retaliation	0.02	0.03	(−0.03 to 0.09)

Notes:
* $p < .05$.

revealed that anthropomorphism was positively associated with perceptions of abuse via perceptions of agency (indirect effect: unstandardized B = 0.14, SE = 0.08, 95% CI = [0.004, 0.32]). These results support Hypothesis 1. (See Table 4.)

We next conducted the same serial mediation analysis to test for Hypothesis 2. We modelled the anthropomorphism condition as the independent variable, agency perceptions as the first mediator, perceptions of abuse as the second mediator, supervisor-directed retaliation as the dependent variable, and experience perception as a control variable. Results revealed a significant serial mediation model (indirect effect: unstandardized B = 0.04, SE = 0.03, 95% CI = [0.001, 0.12]). As a robustness check, we switched the order of the two mediators and results did not support such a serial mediation model (indirect effect: unstandardized B = 0.02, SE = 0.03, 95% CI = [−0.03, 0.09]). These results support Hypothesis 2.

6.3. Discussion

The increased use of robots to take on managerial tasks, such as delivering negative feedback, appears to be an inevitable trend. It is therefore increasingly important to understand how employees react to robot supervisors' actions. To mitigate employees' discomforts with robots in the workplace, scholars and practitioners have advocated for the creation of anthropomorphised, humanoid robots on the grounds that employees will find it more natural to interact with them. Whereas initial evidence has shown anthropomorphism to come with benefits, scholars must also be careful to acknowledge the tactic's potential downsides (Borau et al., 2021; Raisch & Krakowski, 2021; Yam, Bigman, Tang, et al., 2021).

In the present research we proposed that anthropomorphism might be problematic for robots delivering negative feedback. Specifically, we proposed that anthropomorphism would enhance employees' perceptions of the robots' agency, and consequently increase their tendency to perceive the negative feedback as abusive. Two studies offered convergent support for our hypotheses, demonstrating the theorized effects of anthropomorphism and linking them to retaliatory behavior. Importantly, these effects extended to third party observers as well, suggesting that the effects of anthropomorphism in the context of negative feedback are robust and widespread. Below, we discuss the theoretical and practical implications of our work, as well as limitations that point to generative future directions.

6.4. Theoretical contributions

Although pundits have commented on the increased prevalence of robots at work and more specifically how robots would take on more managerial tasks, most discussions of this shift have remained theoretical in nature (e.g., Glikson & Woolley, 2020; Raisch & Krakowski, 2021; Tschang & Mezquita, 2020). Our research thus contributes to a better empirical understanding of this emerging interest in employee-robot interactions. Most importantly, two in-person interactional studies with robots revealed that anthropomorphized robots can actually backfire in the context of receiving negative feedback (Study 1) and observing negative feedback delivered to a colleague (Study 2). This research cautions against the largely unquestioned recommendation to make robots humanlike in appearance (Engle, 2020). More specifically, our work provides insights into the types of tasks where anthropomorphism may backfire – i.e. tasks that elicit negative reactions from employees. This is because when employees have negative interactions with robots, perceptions of humanness carry perceptions of agency and intentionality that lead to more negative reactions to the robots. This then leads employees to retaliate against such robots with greater frequency. In contrast, negative feedback administered by a non-anthropomorphized robot appears to be seen as less agentic and abusive, and thus elicits less retaliation. Consequently, we strive to provide a more balanced perspective which acknowledges how anthropomorphising robots has both benefits and drawbacks depending on the context. Anthropomorphising robots is still a viable choice, just one which we suggest should be considered with care.

More generally, exploring employee-robot interaction necessitates drawing from theories outside of organizational behavior because typical OB theories inherently assume a given level of “human-likeness” of the other party – something not necessarily present when interacting with robots. As such, we introduce mind perception theory to research on employee-robot interaction. Understanding how employees psychologically perceive robots in terms of their agency is an important first

step to examine employees' responses to robots' behaviors. Our theory delineates the psychological mechanisms of the effects – agency perceptions are key to perceptions of abuse and retaliation. Importantly, this is consistent with research on abusive supervision that has discussed the important role of attributions of supervisors' abusive behavior (e.g., Liu, Liao, & Loi, 2012). Rather than focusing on why a supervisor engages in abusive behavior, our work focuses on who is carrying out such behavior and reveals that non-anthropomorphized entities are simply perceived to be unable in carrying out negative behavior because they are perceived to lack the agency to do so.

Finally, our work opens new avenues for future research within the literature on feedback giving. Scholars have predominantly examined how recipients seek out or react to feedback given by human supervisors (e.g., Audia & Locke, 2003; Cannon & Witherspoon, 2005; Chen, Lam, & Zhong, 2007). In contrast, little is known about how employees may intentionally seek out feedback from robot supervisors, the type of feedback they seek, or their reactions to the given feedback. Here, our work suggests that employees can react very differently depending on their perceptions of the robot supervisor's agency. Specifically, non-anthropomorphic robot supervisors were perceived to lack the ability to possess intentions and hence were less likely to be seen as abusive or deserving of retaliation. Through this finding, we suggest that mind perception may be a helpful perspective to adopt in exploring further questions about the when, why, or what kind of feedback employees seek from robot supervisors and how they respond to such feedback.

6.5. Practical contributions

Our work has significant practical implications for those who would like to deploy robots at work. As we alluded to above, anthropomorphized robots might be well-suited to some tasks, but not others. Although it is unrealistic and not economical for organizations to deploy different robots for different tasks, the extent a robot is anthropomorphized can be easily adjusted. Within our experiment, we demonstrate how simply giving robots a human name (i.e., "Paul"), a minimally animated face, and a humanlike voice can cause people to perceive the robot as more "human." This supplements prior research which suggests that simply using a more mechanistic voice can reduce the perceived agency of a robot (Waytz, Heafner, & Epley, 2014), while introducing colloquial slang into the robot's speech content can enhance agency perceptions (Araujo, 2018). These small tweaks can be done between the various managerial tasks enacted by robots to capitalize on the advantages of anthropomorphism – or the lack of it. Despite the ease in which such adjustments could be carried out it is important for organizations to be mindful that the more anthropomorphic the robot, the larger the extent of both the benefits and drawbacks of anthropomorphism. Hence, while organizations can and should still anthropomorphise robots when deemed appropriate, we reiterate the importance of making sure such decisions are carefully and holistically deliberated.

6.6. Limitations and future research

Extant empirical research on human reactions to robots and algorithms has primarily employed scenario-based designs. As such, a strength of our work is experimental realism by having participants actually interact with a robot in a team setting with a confederate. Nonetheless, our experiments created an artificial context in which participants completed tasks that might not have many implications for their identities (unlike employees in actual work contexts), and receiving negative feedback on these tasks might not be as hurtful. Although this suggests that our experiments are conservative, we recommend future research to replicate our work in organizations where employees receive actual feedback from robots.

It is interesting to note that most participants retaliated against the robot supervisor to some extent (3.85 and 3.55 on a 7-point scale in Studies 1 and 2 respectively). However, field studies of abusive supervision and retaliatory response reveal that most employees would not retaliate to human supervisors. For example, in two field studies Lian et al. (2014) found that after being abused by leaders, retaliation is not a common response (1.80 and 1.57 on a 7-point scale). We believe it will be fruitful for future research to examine these discrepancies.[3] In addition, although retaliatory responses to supervisors are significant workplace behaviors (e.g., Bowling & Beehr, 2006; Harvey et al., 2014; Lian et al., 2014), we recommend future research to explore other outcomes. For example, one interesting question for future research is if anthropomorphism influences the extent to which employees act on the feedback they receive. In addition, future research should also examine how anthropomorphism influences third parties' behaviors towards the employees that receive negative feedback, such as the amount of compassion or prosocial behavior these third parties engage in (Priesemuth & Schminke, 2019).

Two additional limitations associated with our manipulation and measurement are important to acknowledge. First, we manipulated anthropomorphism in three ways (name, animated face, and voice). As such, the extent to which these three factors individually impacted perceptions of the robots' humanness is unclear. Another limitation is that although we used a rather well-established measure of agency in our studies, the essence of agency has been described as "the perceived capacity to intend and to act" (Gray et al., 2012, p. 103), and only one of our four items captured the capacity to act. Additionally, our items referred to robots in general rather than the specific robot participants were interacting with. We encourage future research on agency perceptions to address both of these issues with further scale validity work.

Finally, all of our studies were conducted in an Eastern context. We encourage future research to replicate this work in other cultural contexts. Practically, this might be difficult because the use of robots is more common in Asian countries such as China, Japan, and Singapore, relative to the US or Europe. That said, exposure to robots as a result of geographical locations might exert interesting effects on how robots are perceived. For instance, because Americans are generally less exposed to robots, they might find the notion of robots delivering negative feedback as curiosity-inducing and therefore less abusive.

Our research also opens exciting new avenues for research into anthropomorphizing robots. Specifically, given how the same robot (or robots of similar model) can be anthropomorphized or de-anthropomorphized, a natural question which arises is whether and the extent to which individuals will react differently when the same robot (or a similar model) is in a different anthropomorphic "mode." The majority of existing studies, ours included, tend to compare interactions

[3] A related question is whether a human who delivers the same negative feedback would be perceived as worse and more abusive than any robots, anthropomorphized or not. We conducted a pilot study to examine this possibility. Participants from Prolific ($N = 296$) were asked to complete two tasks and then randomly assigned to one of three conditions: two rounds of video-based negative feedback from a human supervisor (a White man), an anthropomorphized robot supervisor, or a non-anthropomorphized supervisor. After receiving the negative feedback, participants completed the same agency ($\alpha = 0.89$) and perceptions of abuse ($\alpha = 0.87$) scales from the extant studies. Based on a sensitivity power analysis with this sample, a statistical power of 80%, and $p < .05$, the minimum effect that can be found is $F = 0.181$. Results revealed that participants rated the human supervisor as highest in agency ($M = 5.67$; $SD = 1.06$), followed by the anthropomorphized robot ($M = 4.29$; $SD = 1.45$), and then the control robot ($M = 3.08$; $SD = 1.61$), $F(2, 295) = 84.17$, $p < .001$. In terms of perceptions of abuse, participants rated the control robot to be the least abusive ($M = 3.68$; $SD = 1.31$), and rated the anthropomorphized robot ($M = 4.13$; $SD = 1.30$) and human supervisor ($M = 4.37$; $SD = 1.24$) to be significantly more abusive, $F(2, 295) = 12.00$, $p < .001$. Interestingly, participants rated the anthropomorphized robot and the human supervisor to be equally abusive ($p = .405$). Future research should thus further examine what managerial tasks are best suited for humans, anthropomorphized robots, or non-anthropomorphized robots.

between distinct anthropomorphised and non-anthropomorphised robots. Consequently, an important next step forward would be to see how human-robot interactions evolve when the same robot (or a similar model) switches between "modes" over time or over specific events. Relatedly, although past research has revealed that anthropomorphism is an effective means to imbue robots with minds, findings on the precise mechanisms of this effect are somewhat mixed. For example, Gray and Wegner (2012) found increases in experience but not agency perceptions, Broadbent et al. (2013) found no differences in either experience and agency perceptions, Müller et al. (2021) and Yam, Bigman, Tang, et al. (2021) found differences in both experience and agency perceptions. Our own studies reveal differences in increased perceptions of agency but not experience. Two possible explanations for these mixed findings include (a) the way in which anthropomorphism is operationalized, and (b) the context in which anthropomorphism is operationalized. For example, in our context of delivering negative feedback participants might have been particularly attuned to the agency aspect of the robots. All in all, we encourage future work to explore what specific facets of anthropomorphism might affect agency and experience perceptions differently.

While people in the industry and marketers appear to see anthropomorphism in mostly positive ways, computer science and especially ethics scholars tend to have a less enthusiastic view. For example, Sharkey and colleagues (Sharkey, 2016; Sharkey & Sharkey, 2021) suggested that anthropomorphising robots in certain industries, such as teaching, might produce more harms than benefits by leading children to develop a false sense of attachment and companionship with robots, in turn impacting their socio-emotional development. Anthropomorphising military robots might also create a false sense of trust, because it implies that a robot can show kindness, mercy, or compassion (Sharkey, 2012). Bryson (2010) went as far as saying that "robots should be slaves" and "should not be described as persons" (p. 1). This is because anthropomorphising robots might mask poor human decision making and the responsibility of robot creators. We hope our work can encourage further exchanges between social psychologists, people in the industry, and ethics scholars in terms of the appropriate use of anthropomorphism in robots.

Our studies revealed that retaliation towards robot supervisors was driven by perceptions of agency. Both legal theory (Foster, 2021) and psychology studies (Gray et al., 2012) link perceptions of agency to being a "moral agent," someone (or something) who is morally responsible for inflicting harm upon others (i.e., abusive supervision) and therefore morally deserving of punishment (i.e., retaliation). But punishment is itself a kind of inflicted harm, and those who suffer harm are typically understood not as "moral agents" but as "moral patients"—someone (or something) vulnerable who feels pain. The ability to feel pain is not tied to agency but instead to experience. Although we focused primarily on agency, future studies should also explore the importance of experience, especially given work on the "harm-made mind" which finds that the very act of inflicting harm increases perceptions of experience (Ward, Olsen, & Wegner, 2013; for harm-made mind research in the robotic contexts, see Küster & Swiderska, 2021; Swiderska & Küster, 2020).

Work on moral typecasting also suggests a potential role of experience. The more we see someone (or something) as a moral agent (capable of inflicting harm on others), the less we see them as a moral patient (vulnerable to suffering pain; Gray & Wegner, 2009). It is possible that that being seen as abusive increases perceptions of the robot's moral agency—and decreases perceptions of moral patiency (i.e., experience), which could counteract any increased perceptions of experience induced by retaliation via the "harm-made mind." Indeed, in the current research we found no impact of our manipulations on perceptions of experience (Study 1 ($t[177] = -0.136$, $p = .892$; Study 2 ($t[162] = -1.613$, $p = .109$), but the current studies were not explicitly designed to test these ideas. Future work should directly explore these interesting questions.

7. Conclusion

With robots taking on increased managerial tasks, it is important to examine how employees react to them and ways to best utilize robots at work. In this paper, we reveal that anthropomorphism – a factor that is found to increase human-robot interaction in many settings – can backfire in the context of delivering negative feedback. Our work thus provides a better theoretical understanding of under what circumstances and why do anthropomorphized robots help vs. hurt organizations.

Open practices

We report all materials used in our studies and these materials can be found in the appendixes. The data for studies 1 and 2 are available at https://osf.io/ug28n/?view_only=913a7f05afac498cb85feddc51cbb72e.

Acknowledgement

This research is supported by a Singapore Ministry of Education Tier 1 Grant (R-317-000-164-115) awarded to the first author.

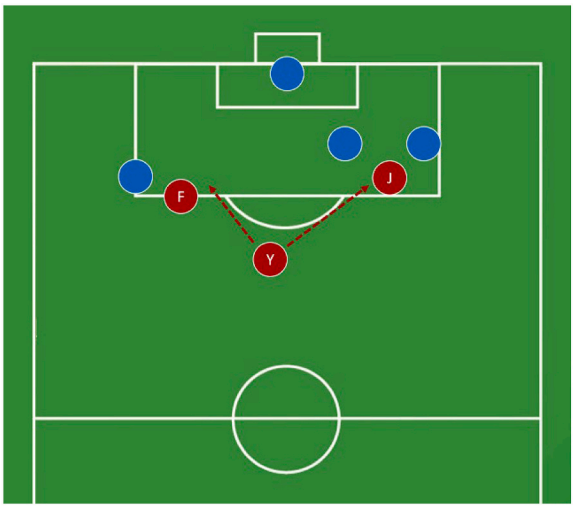
Appendix A. Appendix

A.1. Football knowledge test used in Studies 1–2 (True/False)

1. The winning team of the kick-off coin toss can choose which goal to attack.
2. Players are not allowed to wear multi-coloured/patterned undershirts even if they are the same as the sleeve of the main shirt.
3. Removing one's shirt after scoring a goal can result in a yellow card.
4. Goalkeepers must have at least part of one foot on or in line with the goal-line for penalty kicks.
5. Following a legal kick when either team is awarded a new series, the play clock will be set to 40 s.
6. Each team in football has a maximum of 11 players excluding the goalkeeper.
7. Personal slogans are allowed on player equipment.
8. Referees are in charge of timekeeping the match.
9. Each match consists of two halves of 45 min each, which can be reduced if agreed upon by the referee and both teams.
10. Time taken for penalty kicks count towards the total match time.
11. If a ball touches a match official, it is considered out of play.
12. If a goalkeeper throws the ball directly into the opponent's goal, a goal is awarded.
13. There is a maximum of 5 substitutions of players in any match for each team.
14. Tackling the other team's player gives the other team a direct free kick.
15. An indirect free kick is awarded if the goalkeeper holds the ball for more than 6 s before releasing it within their penalty area.

16. Feinting to take a free kick to confuse opponents is not permitted.
17. A touchline of 105 m is considered valid.
18. Goalposts are 2.44 m in height.
19. If a team kicks a ball into their own goal, it is still considered a goal for the other team.
20. Only the goalkeeper may hold the ball.

A.2. A sample trial of the football simulation used in Studies 1–2



Appendix B. Appendix

B.1. Agency (Studies 1–2)

1. Robots can communicate with others
2. Robots can think
3. Robots can plan their actions
4. Robots can remember things

B.2. Experience (Studies 1–2)

1. Robots can feel pain
2. Robots can feel fear
3. Robots have desires
4. Robots can be happy.

B.3. Perceptions of abuse (Study 1)

This robot…..

1. Ridiculed me
2. Put me down in front of others
3. Reminded me of my past mistakes and failures
4. Didn't give me credit for jobs requiring a lot of effort
5. Made negative comments about me to others
6. Was rude to me
7. Told me that I am incompetent

B.4. Perceptions of abuse (Study 2)

This robot…

1. Ridiculed my partner
2. Put my partner down in front of others
3. Reminded my partner of her past mistakes and failures

4. Didn't give my partner credit for jobs requiring a lot of effort
5. Made negative comments about my partner to others
6. Was rude to my partner
7. Told my partner that she is incompetent

References

Abubshait, A., & Wiese, E. (2017). You look human, but act like a machine: Agent appearance and behavior modulate different aspects of human–robot interaction. *Frontiers in Psychology, 8*. https://doi.org/10.3389/fpsyg.2017.01393

Airenti, G. (2015). The cognitive bases of anthropomorphism: From relatedness to empathy. *International Journal of Social Robotics, 7*(1), 117–127. https://doi.org/10.1007/s12369-014-0263-x

Aquino, K., Tripp, T. M., & Bies, R. J. (2001). How employees respond to personal offense: The effects of blame attribution, victim status, and offender status on revenge and reconciliation in the workplace. *Journal of Applied Psychology, 86*(1), 52–59. https://doi.org/10.1037/0021-9010.86.1.52

Aquino, K., Tripp, T. M., & Bies, R. J. (2006). Getting even or moving on? Power, procedural justice, and types of offense as predictors of revenge, forgiveness, reconciliation, and avoidance in organizations. *Journal of Applied Psychology, 91*, 653–658. https://doi.org/10.1037/0021-9010.92.1.80

Araujo, T. (2018). Living up to the chatbot hype: The influence of anthropomorphic design cues and communicative agency framing on conversational agent and company perceptions. *Computers in Human Behavior, 85*, 183–189. https://doi.org/10.1016/j.chb.2018.03.051

Ashforth, B. E., Schinoff, B. S., & Brickson, S. L. (2020). My Company is Friendly, "Mine'sa Rebel": anthropomorphism and shifting organizational identity from "What" to "Who". *Academy of Management Review, 45*(1), 29–57. https://doi.org/10.5465/amr.2016.0496

Audia, P. G., & Locke, E. A. (2003). Benefiting from negative feedback. *Human Resource Management Review, 13*(4), 631–646. https://doi.org/10.1016/j.hrmr.2003.11.006

Barnes, C. M., Lucianetti, L., Bhave, D. P., & Christian, M. S. (2015). "You wouldn't like me when I'm sleepy": Leaders' sleep, daily abusive supervision, and work unit engagement. *Academy of Management Journal, 58*(5), 1419–1437. https://doi.org/10.5465/amj.2013.1063

Baron, R. A. (1993). Criticism (informal negative feedback) as a source of perceived unfairness in organizations: Effects, mechanisms and countermeasures. In R. Cropanzano (Ed.), *Justice in the workplace* (pp. 115–170). Hillsdale, NJ: Lawrence Erlbaum.

Bartol, K., Durham, C., & Poon, J. (2001). Influence of performance evaluation rating segmentation on motivation and fairness perceptions. *Journal of Applied Psychology, 86*(6), 1106–1119. https://doi.org/10.1037/0021-9010.86.6.1106

Bering, J. M., & Bjorklund, D. F. (2004). The natural emergence of reasoning about the afterlife as a developmental regularity. *Developmental Psychology, 40*(2), 217–233. https://doi.org/10.1037/0012-1649.40.2.217

Bigman, Y. E., & Gray, K. (2018). People are averse to machines making moral decisions. *Cognition, 181*, 21–34. https://doi.org/10.1016/j.cognition.2018.08.003

Bigman, Y. E., Yam, K. C., Marciano, D., Reynolds, S. J., & Gray, K. (2021). Threat of racial and economic inequality increases preference for algorithm decision-making. *Computers in Human Behavior, 122*, Article 106859. https://doi.org/10.1016/j.chb.2021.106859

Bilewicz, M., Imhoff, R., & Drogosz, M. (2011). The humanity of what we eat: Conceptions of human uniqueness among vegetarians and omnivores. *European Journal of Social Psychology, 41*(2), 201–209. https://doi.org/10.1002/ejsp.766

Blair, A., & Saffidine, A. (2019). AI surpasses humans at six-player poker. *Science, 365*(6456), 864–865. https://doi.org/10.1126/science.aay7774

Borau, S., Otterbring, T., Laporte, S., & Fosso Wamba, S. (2021). The most human bot: Female gendering increases humanness perceptions of bots and acceptance of AI. *Psychology & Marketing, 38*(7), 1052–1068. https://doi.org/10.1002/mar.21480

Bowling, N. A., & Beehr, T. A. (2006). Workplace harassment from the victim's perspective: A theoretical model and meta-analysis. *Journal of Applied Psychology, 91*(5), 998–1012. https://doi.org/10.1037/0021-9010.91.5.998

Breazeal, C., Harris, P. L., DeSteno, D., Kory Westlund, J. M., Dickens, L., & Jeong, S. (2016). Young children treat robots as informants. *Topics in Cognitive Science, 8*(2), 481–491. https://doi.org/10.1111/tops.12192

Brink, K. A., Gray, K., & Wellman, H. M. (2019). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Development, 90*(4), 1202–1214. https://doi.org/10.1111/cdev.12999

Broadbent, E., Kumar, V., Li, X., Sollers, J., 3rd, Stafford, R. Q., MacDonald, B. A., & Wegner, D. M. (2013). Robots with display screens: A robot with a more humanlike face display is perceived to have more mind and a better personality. *PLoS One, 8*(8), Article e72589. https://doi.org/10.1371/journal.pone.0072589

Brown, B. R. (1968). The effects of need to maintain face on interpersonal bargaining. *Journal of Experimental Social Psychology, 4*(1), 107–122. https://doi.org/10.1016/0022-1031(68)90053-x

Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. WW Norton & Company.

Bryson, J. J. (2010). Robots should be slaves. In , *Vol. 8. Close engagements with artificial companions: Key social, psychological, ethical and design issues* (pp. 63–74). https://doi.org/10.1075/nlp.8.11bry

Cairns, R. (2021 Aug 26). More than 50 robots are working at Singapore's high-tech hospital. https://edition.cnn.com/2021/08/25/asia/cgh-robots-healthcare-spc-intl-hnk/index.html.

Cannon, M. D., & Witherspoon, R. (2005). Actionable feedback: Unlocking the power of learning and performance improvement. *Academy of Management Perspectives, 19*(2), 120–134. https://doi.org/10.5465/ame.2005.16965107

Carey, S. (1985). *Conceptual change in childhood*. MIT press.

Castelo, N., & Ward, A. F. (2021). Conservatism predicts aversion to consequential artificial intelligence. *PLoS One, 16*(12), Article e0261467.

Chamorro-Premuzic, T., & Ahmetoglu, G. (2016). The pros and cons of robot managers. *Harvard Business Review, 2016*(12).

Chan, M. E., & McAllister, D. J. (2014). Abusive supervision through the lens of employee state paranoia. *Academy of Management Review, 39*(1), 44–66. https://doi.org/10.5465/amr.2011.0419

Chan, W., & Simester, A. (2011). Four functions of *Mens Rea*. *The Cambridge Law Journal, 70*, 381–396.

Chen, Z., Lam, W., & Zhong, J. A. (2007). Leader-member exchange and member performance: A new look at individual-level negative feedback-seeking behavior and team-level empowerment climate. *Journal of Applied Psychology, 92*(1), 202–212. https://doi.org/10.1037/0021-9010.92.1.202

Cohen, C., & Regan, T. (2001). *The animal rights debate*. Rowman & Littlefield.

Cuddy, A. J., Fiske, S. T., & Glick, P. (2008). Warmth and competence as universal dimensions of social perception: The stereotype content model and the BIAS map. *Advances in Experimental Social Psychology, 40*, 61–149. https://doi.org/10.1016/s0065-2601(07)00002-0

Dietvorst, B. J., Simmons, J. P., & Massey, C. (2015). Algorithm aversion: People erroneously avoid algorithms after seeing them err. *Journal of Experimental Psychology: General, 144*(1), 114–126. https://doi.org/10.1037/xge0000033

Dix, T., Ruble, D. N., Grusec, J. E., & Nixon, S. (1986). Social cognition in parents: Inferential and affective reactions to children of three age levels. *Child Development, 57*(4), 879–894. https://doi.org/10.2307/1130365

Dubow, S. (2010). *Ourselves unborn: A history of the fetus in modern America*. Oxford University Press.

Dupré, K. E., Inness, M., Connelly, C. E., Barling, J., & Hoption, C. (2006). Workplace aggression in teenage part-time employees. *Journal of Applied Psychology, 91*(5), 987–997. https://doi.org/10.1037/0021-9010.91.5.987

Engle, J. (2020, March 2). Lesson of the day: 'Should robots have a face?'. *The New York Times*. https://www.nytimes.com/.

Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition, 26*(2), 143–155. https://doi.org/10.1521/soco.2008.26.2.143

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295x.114.4.864

Eschleman, K. J., Bowling, N. A., Michel, J. S., & Burns, G. N. (2014). Perceived intent of supervisor as a moderator of the relationships between abusive supervision and counterproductive work behaviours. *Work and Stress, 28*(4), 362–375. https://doi.org/10.1080/02678373.2014.961183

Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology, 51*(4), 724–731. https://doi.org/10.1111/j.2044-8309.2011.02082.x

Fehr, E., & Fischbacher, U. (2004). Third-party punishment and social norms. *Evolution and Human Behavior, 25*(2), 63–87. https://doi.org/10.1016/s1090-5138(04)00005-4

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature, 415*(6868), 137–140. https://doi.org/10.1038/415137a

Fehr, R., Gelfand, M. J., & Nag, M. (2010). The road to forgiveness: A meta-analytic synthesis of its situational and dispositional correlates. *Psychological Bulletin, 136*(5), 894. https://doi.org/10.1037/a0019993

Ferris, G. R., Bhawuk, D. P., Fedor, D. F., & Judge, T. A. (2018). Organizational politics and citizenship: Attributions of intentionality and construct definition. In *Attribution theory* (pp. 231–252). Routledge.

Finkelstein, S. R., & Fishbach, A. (2012). Tell me what I did wrong: Experts seek and respond to negative feedback. *Journal of Consumer Research, 39*(1), 22–38. https://doi.org/10.1086/661934

Foster. (2021, July 30). Mens Rea: An overview of state-of-mind requirements for federal criminal offenses. *Congressional Research Service*. https://crsreports.congress.gov/product/pdf/R/R46836/1.

Fraune, M. R. (2020). Our robots, our team: robot anthropomorphism moderates group effects in human–robot teams. *Frontiers in Psychology, 11*. https://doi.org/10.3389/fpsyg.2020.01275

Glikson, E., & Woolley, A. W. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627–660. https://doi.org/10.5465/annals.2018.0057

Gombolay, M. C., Gutierrez, R. A., Clarke, S. G., Sturla, G. F., & Shah, J. A. (2015). Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Autonomous Robots, 39*(3), 293–312. https://doi.org/10.1007/s10514-015-9457-9

Granulo, A., Fuchs, C., & Puntoni, S. (2019). Psychological reactions to human versus robotic job replacement. *Nature Human Behaviour, 3*(10), 1062–1069. https://doi.org/10.1038/s41562-019-0670-y

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812), 619. https://doi.org/10.1126/science.1134475

Gray, K., Jenkins, A. C., Heberlein, A. S., & Wegner, D. M. (2011). Distortions of mind perception in psychopathology. *Proceedings of the National Academy of Sciences, 108*(2), 477–479. https://doi.org/10.1073/pnas.1015493108

Gray, K., & Wegner, D. M. (2008). The sting of intentional pain. *Psychological Science, 19*(12), 1260–1262. https://doi.org/10.1111/j.1467-9280.2008.02208.x

Gray, K., & Wegner, D. M. (2009). Moral typecasting: Divergent perceptions of moral agents and moral patients. *Journal of Personality and Social Psychology, 96*(3), 505–520. https://doi.org/10.1037/a0013748

Gray, K., & Wegner, D. M. (2012). Feeling robots and human zombies: Mind perception and the uncanny valley. *Cognition, 125*(1), 125–130. https://doi.org/10.1016/j.cognition.2012.06.007

Gray, K., Young, L., & Waytz, A. (2012). Mind perception is the essence of morality. *Psychological Inquiry, 23*(2), 101–124. https://doi.org/10.1080/1047840x.2012.651387

Harvey, P., Harris, K. J., Gillis, W. E., & Martinko, M. J. (2014). Abusive supervision and the entitled employee. *The Leadership Quarterly, 25*(2), 204–217. https://doi.org/10.1016/j.leaqua.2013.08.001

Hayes, A. F. (2017). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

Hershcovis, M. S., & Barling, J. (2010). Towards a multi-foci approach to workplace aggression: A meta-analytic review of outcomes from different perpetrators. *Journal of Organizational Behavior, 31*(1), 24–44. https://doi.org/10.1002/job.621

IFR. (2022). Topics and definitions. https://ifr.org/.

Kant, I. (1964). *Groundwork for the metaphysics of morals*. New York: Harper & Row.

Kruse, K. (2018, November 28). In 2023 your boss will be a robot (and you will love her). *Forbes*. https://www.forbes.com.

Küster, D., & Swiderska, A. (2021). Seeing the mind of robots: Harm augments mind perception but benevolent intentions reduce dehumanisation of artificial entities in visual vignettes. *International Journal of Psychology, 56*(3), 454–465. https://doi.org/10.1002/ijop.12715

Lian, H., Brown, D. J., Ferris, D. L., Liang, L. H., Keeping, L. M., & Morrison, R. (2014). Abusive supervision and retaliation: A self-control framework. *Academy of Management Journal, 57*(1), 116–139. https://doi.org/10.5465/amj.2011.0977

Liu, D., Liao, H., & Loi, R. (2012). The dark side of leadership: A three-level investigation of the cascading effect of abusive supervision on employee creativity. *Academy of Management Journal, 55*(5), 1187–1212. https://doi.org/10.5465/amj.2010.0400

Malle, B. F., Guglielmo, S., & Monroe, A. E. (2014). A theory of blame. *Psychological Inquiry, 25*(2), 147–186. https://doi.org/10.1080/1047840x.2014.877340

Martin, R., Guillaume, Y., Thomas, G., Lee, A., & Epitropaki, O. (2015). Leader-member exchange (LMX) and performance: A Meta-analytic review. *Personnel Psychology, 69*(1), 67–121. https://doi.org/10.1111/peps.12100

McAuliffe, K., Jordan, J. J., & Warneken, F. (2015). Costly third-party punishment in young children. *Cognition, 134*, 1–10. https://doi.org/10.1016/j.cognition.2014.08.01

Miketta, S., & Friese, M. (2019). Debriefed but still troubled? About the (in) effectiveness of postexperimental debriefings after ego threat. *Journal of Personality and Social Psychology, 117*(2), 282–309. https://doi.org/10.1037/pspa0000155

Mikołajczak, M., & Bilewicz, M. (2015). Foetus or child? Abortion discourse and attributions of humanness. *British Journal of Social Psychology, 54*(3), 500–518. https://doi.org/10.1111/bjso.12096

Mitchell, M. S., & Ambrose, M. L. (2007). Abusive supervision and workplace deviance and the moderating effects of negative reciprocity beliefs. *Journal of Applied Psychology, 92*(4), 1159–1168. https://doi.org/10.1037/0021-9010.92.4.1159

Mitchell, M. S., Vogel, R. M., & Folger, R. (2015). Third parties' reactions to the abusive supervision of coworkers. *Journal of Applied Psychology, 100*(4), 1040. https://doi.org/10.1037/apl0000002

Müller, B. C., Gao, X., Nijssen, S. R., & Damen, T. G. (2021). I, robot: How human appearance and mind attribution relate to the perceived danger of robots. *International Journal of Social Robotics, 13*(4), 691–701.

Murphy, M. (2017). This bricklaying robot is going to destroy construction jobs. *New York Post, 27*. Mar https://nypost.com/2017/03/27/this-bricklaying-robot-is-going-to-destroy-construction-jobs.

Newman, D. T., Fast, N. J., & Harmon, D. J. (2020). When eliminating bias isn't fair: Algorithmic reductionism and procedural justice in human resource decisions. *Organizational Behavior and Human Decision Processes, 160*, 149–167. https://doi.org/10.1016/j.obhdp.2020.03.008

Oracle. (2019, October 15). New study: 64% of people trust a robot more than their manager. https://www.oracle.com/corporate/pressrelease/robots-at-work-101519.html.

Priesemuth, M., & Schminke, M. (2019). Helping thy neighbor? Prosocial reactions to observed abusive supervision in the workplace. *Journal of Management, 45*(3), 1225–1251. https://doi.org/10.1177/0149206317702219

Rai, T. S., & Diermeier, D. (2015). Corporations are cyborgs: Organizations elicit anger but not sympathy when they can think but cannot feel. *Organizational Behavior and Human Decision Processes, 126*, 18–26. https://doi.org/10.1016/j.obhdp.2014.10.001

Raisch, S., & Krakowski, S. (2021). Artificial intelligence and management: The automation–augmentation paradox. *Academy of Management Review, 46*(1), 192–210. https://doi.org/10.5465/amr.2018.0072

Raver, J. L., Jensen, J. M., Lee, J., & O'Reilly, J. (2012). Destructive criticism revisited: Appraisals, task outcomes, and the moderating role of competitiveness. *Applied Psychology, 61*(2), 177–203. https://doi.org/10.1111/j.1464-0597.2011.00462.x

Roesler, E., Onnasch, L., & Majer, J. I. (2020). The effect of anthropomorphism and failure comprehensibility on human-robot trust. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 64*(1), 107–111. https://doi.org/10.1177/1071181320641028

Sayre, F. B. (1932). Mens Rea. *Harvard Law Review, 45*(6), 974–1026. https://doi.org/10.2307/1332142

Schönfeld, M. (1992). Who or what has moral standing? *American Philosophical Quarterly, 29*(4), 353–362. https://doi.org/10.1111/1467-9213.00105

Schyns, B., Felfe, J., & Schilling, J. (2018). Is it me or you?—How reactions to abusive supervision are shaped by leader behavior and follower perceptions. *Frontiers in Psychology, 9*, 1309. https://doi.org/10.3389/fpsyg.2018.01309

Scott, T. (2021, July 7). *Difference between algorithm and artificial intelligence*. Data Science Central. https://www.datasciencecentral.com/difference-between-algorithm-and-artificial-intelligence/.

Sharkey, A., & Sharkey, N. (2021). We need to talk about deception in social robotics! *Ethics and Information Technology, 23*(3), 309–316. https://doi.org/10.1007/s10676-020-09573-9

Sharkey, A. J. (2016). Should we welcome robot teachers? *Ethics and Information Technology, 18*(4), 283–297. https://doi.org/10.1007/s10676-016-9387-z

Sharkey, N. E. (2012). The evitability of autonomous robot warfare. *International Review of the Red Cross, 94*(886), 787–799. https://doi.org/10.1017/s1816383112000732

Sheridan, T. B. (2016). Human–robot interaction: Status and challenges. *Human Factors, 58*(4), 525–532. https://doi.org/10.1177/0018720816644364

Skarlicki, D. P., & Folger, R. (1997). Retaliation in the workplace: The role of distributive, procedural, and interactional justice. *Journal of Applied Psychology, 82*(3), 434–443. https://doi.org/10.1037/0021-9010.82.3.434

Skowronski, J. J., & Carlston, D. E. (1987). Social judgment and social memory: The role of cue diagnosticity in negativity, positivity, and extremity biases. *Journal of Personality and Social Psychology, 52*(4), 689–699. https://doi.org/10.1037/0022-3514.52.4.689

Soper, S. (2021, June 28). Fired by bot at Amazon: 'It's you against the machine'. *Bloomberg*. https://www.bloomberg.com.

Stearns, P. N. (2020). *The industrial revolution in world history*. Routledge.

Strait, M., Vujovic, L., Floerke, V., Scheutz, M., & Urry, H. (2015, April). Too much humanness for human-robot interaction: Exposure to highly humanlike robots elicits aversive responding in observers. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. https://doi.org/10.1145/2702123.2702415

Swiderska, A., & Küster, D. (2020). Robots as malevolent moral agents: Harmful behavior results in dehumanization, not anthropomorphism. *Cognitive Science, 44*(7). https://doi.org/10.1111/cogs.12872

Tang, S., & Gray, K. (2018). CEOs imbue organizations with feelings, increasing punishment satisfaction and apology effectiveness. *Journal of Experimental Social Psychology, 79*, 115–125. https://doi.org/10.1016/j.jesp.2018.06.002

Tepper, B. J. (2000). Consequences of abusive supervision. *Academy of Management Journal, 43*(2), 178–190. https://doi.org/10.2307/1556375

Tesser, A., & Rosen, S. (1975). The reluctance to transmit bad news. In *, Vol. 8. Advances in experimental social psychology* (pp. 193–232). https://doi.org/10.1016/s0065-2601(08)60251-8

Tschang, F. T., & Mezquita, E. A. (2020). Artificial intelligence as augmenting automation: Implications for employment. *Academy of Management Perspectives*. https://doi.org/10.5465/amp.2019.0062

Vollmer, A. L., Read, R., Trippas, D., & Belpaeme, T. (2018). Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Science. Robotics, 3*(21). https://doi.org/10.1126/scirobotics.aat7111

Wada, K., & Shibata, T. (2007). Living with seal robots—Its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics, 23*(5), 972–980. https://doi.org/10.1109/tro.2007.906261

Wang, H., Law, K. S., Hackett, R. D., Wang, D., & Chen, Z. X. (2005). Leader-member exchange as a mediator of the relationship between transformational leadership and followers' performance and organizational citizenship behavior. *Academy of Management Journal, 48*(3), 420–432. https://doi.org/10.5465/amj.2005.17407908

Ward, A. F., Olsen, A. S., & Wegner, D. M. (2013). The harm-made mind: Observing victimization augments attribution of minds to vegetative patients, robots, and the dead. *Psychological Science, 24*(8), 1437–1445. https://doi.org/10.1177/0956797612472343

Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science, 5*(3), 219–232. https://doi.org/10.1177/1745691610369336

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

Yam, K. C., Bigman, Y., & Gray, K. (2021). Reducing the uncanny valley by dehumanizing humanoid robots. *Computers in Human Behavior, 125*, Article 106945. https://doi.org/10.1016/j.chb.2021.106945

Yam, K. C., Bigman, Y., Tang, P. M., Ilies, R., De Cremer, D., Soh, H., & Gray, K. (2021). Robots at work: People prefer—And forgive—Service robots with perceived feelings. *Journal of Applied Psychology, 106*(10), 1557–1572. https://doi.org/10.1037/apl0000834

Young, J., & Cormier, D. (2014, April 2). Can robots be managers, too? *Harvard Business Review*. https://hbr.org/.

Zhang, T., Kaber, D. B., Zhu, B., Swangnetr, M., Mosaly, P., & Hodge, L. (2010). Service robot feature design effects on user perceptions and emotional responses. *Intelligent Service Robotics, 3*(2), 73–88. https://doi.org/10.1007/s11370-010-0060-9