

Service robots and artificial morality: an examination of robot behavior that violates human privacy

Magnus Söderlund

Centre for Consumer Marketing, Stockholm School of Economics, Stockholm, Sweden

Abstract

Purpose – Service robots are expected to become increasingly common, but the ways in which they can move around in an environment with humans, collect and store data about humans and share such data produce a potential for privacy violations. In human-to-human contexts, such violations are transgression of norms to which humans typically react negatively. This study examines if similar reactions occur when the transgressor is a robot. The main dependent variable was the overall evaluation of the robot.

Design/methodology/approach – Service robot privacy violations were manipulated in a between-subjects experiment in which a human user interacted with an embodied humanoid robot in an office environment.

Findings – The results show that the robot's violations of human privacy attenuated the overall evaluation of the robot and that this effect was sequentially mediated by perceived robot morality and perceived robot humanness. Given that a similar reaction pattern would be expected when humans violate other humans' privacy, the present study offers evidence in support of the notion that humanlike non-humans can elicit responses similar to those elicited by real humans.

Practical implications – The results imply that designers of service robots and managers in firms using such robots for providing service to employees should be concerned with restricting the potential for robots' privacy violation activities if the goal is to increase the acceptance of service robots in the habitat of humans.

Originality/value – To date, few empirical studies have examined reactions to service robots that violate privacy norms.

Keywords Automated services, Service robots, Artificial morality, Privacy violations

Paper type Research paper

1. Introduction

Humanlike service robots are expected to become more common (Belanche *et al.*, 2020; Lu *et al.*, 2020; Swiderska and Küster, 2020; Sullins, 2011). As the level of sophistication of such robots increases, they are also likely to become involved in tasks with moral implications (Misselhorn, 2018; Moor, 2006; Sullins, 2006). And when such implications do exist, many observers have noted that robot inability to reason and/or act morally will make humans uncomfortable, or even harmed; thus, it has been recommended that robots should be provided with morality (Allen *et al.*, 2006; Crnkovic and Çürüklü, 2012; Malle, 2014; Moor, 2006; Scheutz, 2017; Sullins, 2011).

Morality can be seen as the capability to distinguish what is right from what is wrong and trying to do what is right (Gray *et al.*, 2012). The extent to which current artificial agents can – or will be able to – have morality in this way is indeed subject to debate (e.g. Coeckelbergh, 2010; Misselhorn, 2018; Moor, 2006; Scheutz, 2017; Sullins, 2006). However, given that



we humans typically apply a scheme for human-to-human interactions when we interact with humanlike non-human agents and that we imbue such agents with humanlike attributes (Epley, 2018; Reeves and Nass, 1996), robots can be *perceived* as if they have morality (Banks, 2019; Coeckelbergh, 2010). Perceptions of this type are the subject of the present study. More specifically, as in signaling theory, it is assumed that robotic behavior with moral implications provides an observer with cues about robot morality (an inherently unobservable characteristic *per se*) in the same way as, say, the frequency of advertising spending and advertising repetition can provide the receiver of ads with cues about product quality (Kirmani, 1997; Moorthy and Zhao, 2000).

Many behaviors of a service robot can influence perceptions of its morality, and the present study examines one specific behavioral category: how a service robot deals with human privacy. As noted by Calo (2012), and practically by definition, robots are equipped with the abilities to sense, process and record the world around them, which can lead to many robot behaviors that violate the privacy of humans (Calo, 2012; Sullins, 2011; Syrdal *et al.*, 2007). This should be seen in the light of advanced service robots being expensive. Therefore, they are likely to be shared between several users (Lee *et al.*, 2011) and sharing gives robots opportunities to collect information about several humans (and to share this information with several humans). It may also be noted that robots can acquire a wider range of information about humans compared to other means that have been discussed previously in privacy terms – such as loyalty programs, internet sites and CCTV (Syrdal *et al.*, 2007). Hence it has been argued that robots should be programmed to be “privacy-sensitive” and, more specifically, that their behavior should be constrained so that they are explicitly instructed about what they should not do (Rueben and Smart, 2016). To accomplish this, however, when robots are powered by artificial intelligence, which may result in autonomous learning in such a way that not even the robot itself knows how it makes decisions (Ishii, 2019), appears to be a paramount challenge.

Service robots of the type covered by the present study have not been around long enough to generate a list of concrete examples of actual privacy violations. Privacy-related concerns, however, have already been raised when it comes to less sophisticated robots types. Robot vacuum cleaners, for example, create maps of the homes that they clean and those maps may comprise information about the existence of a child in the household and what room this child is in. Exactly how these data are used is unclear, but the mere possibility that they can be sold to third parties has created uneasiness (Astor, 2017). Similarly, for smart voice assistants, such as Amazon’s Alexa, which are already in the homes of millions of users, many privacy-related incidents have been reported; for example, when the assistant wakes up unprompted or sends recordings from one person’s home to other people (Lynskey, 2019). Moreover, service robots are likely to share some of the technology (e.g. many sensors and internet connection) used in most new cars – and this indeed raises privacy concerns. For example, when a reporter from *The Washington Post* hacked the computer in his car, he found that it contained not only maps of where he had been driving but also much of the content of his cell phone (Fowler, 2019).

In any event, privacy issues in human-to-robot interactions have been under-researched, and several authors have called for more research on the privacy challenges of social robots (e.g. Calo, 2012; Krupp *et al.*, 2017; Rueben *et al.*, 2018). Some studies have indeed been made, and they do indicate that the behavior of social robots can affect observers’ views of the extent to which robots respect humans’ privacy (e.g. Hedao *et al.*, 2019). Other studies, however, indicate that humans are naïve with respect to what data robots can collect and disseminate (Lee *et al.*, 2011). This seems to mirror a general lack of awareness among individuals regarding what information other people, firms and governments have about them and how that information is used (Acquisti *et al.*, 2015). What is clear from existing research on other services than those that are robot-related, however, is this: privacy concerns

stemming from one particular service have a negative impact on people’s usage of the service (Baruh *et al.*, 2017). Ultimately, then, if service robots are to become widely used, people’s perceptions of the privacy-related consequences of the behavior of such robots must be addressed.

The purpose of the present study is to examine if service robot behavior comprising sharing of human personal data affects perceptions of robot morality, perceived robot humanness and the overall evaluation of the robot. The latter response was chosen as the main downstream variable because (in human-to-human service contexts) the service provider typically *is* the service from the customer’s point of view (Bitner *et al.*, 1990). Thus the evaluation of the individual service provider is likely to be a main determinant of the overall evaluation of the service itself and of the firm that the provider represents. To this end, an experiment was used to manipulate the sharing-of-information behavior of an embodied service robot that interacted with a human. The specific setting in the experiment comprised a situation in which a humanlike robot – capable of collecting, storing and retrieving human conversations – was asked by a human to disclose information about a non-present human.

2. Theoretical framework and hypotheses

2.1 Main thesis and conceptual point of departure

The main thesis in the present study is that when a service robot shares personal information so that the individual from whom information has been collected is subject to loss of control of information about him/herself (i.e. there is a violation of the individual’s privacy), the robot will be subject to (1) attenuated morality perceptions. Such perceptions, in turn, are assumed to be (2) positively associated with perceptions of the robot’s humanness, and (3) humanness perceptions are assumed to be positively associated with the overall evaluation of the robot (see Figure 1). The expected net outcome is that the robot is subject to less positive evaluations when it engages in privacy-violating behavior compared to privacy-preserving behavior. The arguments behind these assumptions are discussed in the sections below.

The conceptual point of departure is that we humans tend to react to (humanlike) non-humans in ways that are similar to how we react to real humans. This reaction pattern is typically referred to as anthropomorphism (Epley *et al.*, 2008; Epley, 2018). One main reason why such reactions are likely is that we humans are equipped with evolution-based social responses in relation to other humans – responses that occur more or less automatically in interaction situations resembling the situations for which they were originally developed (Nowak and Biocca, 2003). That is to say, because of the associative nature of human brains, exposure to a non-human object that is similar in some ways to a real human can make accessible and activate mental content related to real humans, and in the next step this content is applied, more or less automatically, to the non-human object (Epley, 2018). A main idea in the robotic literature is that robots used for interactions with humans should be designed so that they encourage anthropomorphism (e.g. by giving them a voice and a face), because this is likely to enhance acceptability (Damiano and Dumouchel, 2018). For studies of humans’ reactions to (somewhat humanlike) service robots, this means that theory derived from the context of human-to-human interactions can be useful, and it is in this way the present study’s hypotheses have been developed.

Figure 1.
Privacy-violating
behavior of a service
robot and the impact on
the overall evaluation
of the robot



Source(s): Figure created by author

2.2 Privacy and privacy violations

Privacy in the present study has to do with being in control of information about oneself (Margulis, 2003; Moore, 2003; Nissenbaum, 2004; Rueben and Smart, 2016; Smith *et al.*, 2011). That is to say, if one cannot control others' access to self-related information, a low level of privacy is at hand; if one can fully control others' access to self-related information, a high level of privacy is maintained (Derlega and Chaikin, 1977). With this view, then, a high level of privacy for an individual exists when he or she has voluntarily chosen what self-information others have access to (Moore, 2003). It may be noted that a control-based view of privacy is not only an academic attempt to define a theoretical construct; it is the basis for several regulations such as the European Union's General Data Protection Regulation (GDPR). Moreover, given that a high level of control, in general, is more positively valenced than a low level of control (White, 1959), the control aspect *per se* seems to be part of the reason why privacy violations typically produce a negative state of mind for the victim.

A high level of privacy, however, is assumed to produce many other benefits than a positive state of mind. For the individual, it has been argued that a zone of "relative insularity" – in which an individual is free to experiment with formulating preferences, goals, values and various conceptions of self – can boost the individual's autonomy, thoughtfulness, personal growth and maturity (Magi, 2011; Moore, 2003; Margulis, 2003; Nissenbaum, 2004; Rueben and Smart, 2016). Privacy is also likely to protect the individual from being exploited by others (Derlega and Chaikin, 1977; Magi, 2011; Smith *et al.*, 2011) and several authors have stressed the importance of privacy as a precondition for the formation of intimate relationships and friendship (Magi, 2011; Rueben and Smart, 2016). Moreover, at the society level, it has been argued that privacy promotes a vital democracy (Magi, 2011; Margulis, 2003; Nissenbaum, 2004) and that it may serve as an insurance policy against the emergence of totalitarianism (Nissenbaum, 2004). It has even been claimed that populations that fail to achieve a minimum level of privacy for individual members would self-destruct in various ways (Moore, 2003).

Given that privacy is valuable from several points of view and is seen by participants in empirical studies as highly important to uphold in human-to-human relationships (Argyle *et al.*, 1985), it is not surprising that numerous standards have been developed to protect the individual's privacy. They comprise social norms for everyday behavior, policies, codes of conduct, declaration of rights, soft laws and laws. It even seems as if privacy is built into the very fabric of social establishments, in the sense that doors, fences, window blinds and walls are omnipresent in human environments (Moore, 2003). All this, then, imposes restrictions on privacy-threatening activities such as surveillance of people, access to private information about them, intrusions into their private places and disclosure of personal facts (Nissenbaum, 2004; Rueben and Smart, 2016). It is also in the light of such norms that the notion of privacy violation should be seen: a privacy violation is a transgression of norms (Nissenbaum, 2004).

There are several ways in which service robots may violate humans' privacy. For example, when robots are engaged in surveillance and record and transmit data about individuals, such as the location of a person and what a person says; when they have access to spaces in which sensitive information is available; and when they extract information in their direct interactions with a user (Calo, 2012; Rueben and Smart, 2016). Specific examples of robot activities seen as problematic from a privacy point of view, by participants in studies, are if robots get access to bedrooms and bathrooms, if they catch people in a state of undress or in embarrassing situations, if others are able to hack the robot and get access to its data and if data from the robot somehow end up in the hands of marketers (Krupp *et al.*, 2017).

In the present study, the focal privacy-violation activity is robots' sharing of personal information in such a way that it represents a loss of control for the person from whom information has been collected. Given the importance humans attach to privacy in human-to-human settings, as reflected by the wide array of norms that serve to protect privacy,

particularly a social norm that information received in confidence from one person should not be disclosed to others (Argyle *et al.*, 1985) and given anthropomorphism (Epley, 2018; Reeves and Nass, 1996), it is expected in the present study that we humans are able to perceive acts by a service robot as violations of privacy if such acts appear to reduce a human's control of information about him/herself.

2.3 Privacy-violating behavior and morality perceptions

As already indicated, the violation of privacy can be seen as a transgression of norms (Nissenbaum, 2004) and, in general, we humans are sensitive to when others do not behave according to norms. Indeed, it seems as if we are hardwired to quickly assess if a person is a norm violator (van Kleef *et al.*, 2015; van Leeuwen *et al.*, 2012; Malle, 2014). Empirically, it has been shown, for example, that statements in conflict with one's moral values elicit extremely rapid neural responses (Van Berkum *et al.*, 2009). However, there are different views in the literature on what happens once a violation of norms has been detected. Yet, several theories assume that one more or less immediate response is moral judgment (i.e. an assessment of the extent to which an act is wrong or right). Then, such judgments are expected to be employed as a source of inferences about a violator's agency and the violator's beliefs about the act's potential to cause harm (Guglielmo, 2015). In other words, with this view, moral judgment is seen as a quick, intuitive response, which is followed by slow, ex post moral reasoning about the agent (Haidt, 2001). Examples of results of such reasoning is that intentional violators are blamed more than those that are unintentional (Malle *et al.*, 2014), that harm caused by agents who are seen as malicious is perceived as more painful than accidental harm (Gray *et al.*, 2014), that violators who appear to not feel any guilt are perceived as highly immoral (Harenski and Kiehl, 2011) and that children, who appear less autonomous than adults, are held less responsible for transgressions (Bigman *et al.*, 2019).

In the present study, it is assumed that moral reasoning elicited by a norm violation is one basis for attributions of mind to a person. One central aspect of having mind is morality, which is about the extent to which a person has the capability to distinguish between what is right from what is wrong and the capability to try to do what is right (Gray *et al.*, 2012). Thus an assessment of someone's morality is an assessment of a capability, not an assessment if one particular act is right or wrong. It is also assumed that there is a strong imperative among humans to actually *use* the capability to distinguish between what is right and wrong – so that right rather than wrong activities are chosen. That is to say, we humans are expected to do what is good, not what is bad, in inter-human relations (Baumeister *et al.*, 2001). For an observer, who has no access to the detailed content of others' minds, but can use observations of another person's behavior as clues about his or her mind, it is further assumed that the valence of an act in relation to the imperative to do what is good is used for inferences about the level of the person's morality – in such a way that a human target person who does what is right is ascribed more morality than a person who does what is wrong. More specifically, it is expected that a human norm violator will be perceived as lower in morality than a human preserver of norms when it comes to acts that affect other persons' privacy.

Given anthropomorphism, and given that humans can be attributed different levels of morality, one would expect that artificial agents with humanlike characteristics can be imbued with different levels of morality, too. Empirically, the results in Malle *et al.* (2016) indicate that most of their participants accepted that robots can be subject to moral judgment, and the participants did not attribute less blame to humanoid robotic agents in relation to human agents in moral dilemma situations. This thus indicates that these robots were attributed at least some level of morality. Moreover, the results in Weisman *et al.* (2017) indicate that robots are perceived as having a low level (but not a zero level) of the ability to distinguish between what is right and wrong. Given that humans who engage in

norm-violating behavior are attributed lower morality compared to those who are norm-compliant, the same pattern of attribution is expected for humanlike non-human agents. Therefore, the following is hypothesized for a situation in which a service robot shares personal information about a human in such a way that the human's privacy is violated:

- H1.* Privacy-violating behavior by a robot is negatively associated with perceived robot morality.

2.4 Perceived morality and perceived humanness

Perceived humanness has to do with the extent to which an individual is seen as having characteristics that are typical for humans. This has been referred to as the “human nature” aspect of humanness (Haslam, 2006). As a perceptual dimension, both real humans (Epley *et al.*, 2013; Söderlund, 2020) and non-humans (Bastian *et al.*, 2012; Kim and Sundar, 2012; Powers and Kiesler, 2006) have been assessed in terms of perceived humanness. This variable has been included in the present study because it is frequently assumed that the perceived humanness of an artificial agent boosts humans' acceptance and enhances the comfort with which they interact with it (Epley, 2018; Kahn *et al.*, 2007).

Given that morality is a fundamental part of what it means to be a human (Goodwin *et al.*, 2014; Mikhail, 2007), it is expected that perceived morality can boost perceived humanness. Conversely, humans who are seen as lacking morality can be dehumanized (Bastian *et al.*, 2013). Indirect evidence for this is provided also by Söderlund (2020), a study showing that norm-violating humans were perceived as lower in humanness than norm-compliant humans. For non-human agents, Söderlund and Oikarinen (2021) have identified a positive association between perceived morality and perceived humanness for virtual agents used by firms in consumer settings. Similar results appear in Banks (2019). The following is therefore hypothesized for the situation in which a service robot engages in privacy-violating behavior by sharing personal information about a human:

- H2.* Perceived robot morality is positively associated with perceived robot humanness.

2.5 Perceived humanness and overall evaluations

Furthermore, it is expected that the perceived humanness of an agent has a positive influence on the overall evaluation of the agent. One main reason is that humans in general have an overall positive rather than a negative charge (Sears, 1983). This is likely to be a function of the inherently social nature of humans; we humans need other humans for both practical and existential issues (Epley *et al.*, 2008). From an evolutionary point of view, then, it seems reasonable that humans are hardwired to evaluate what is human in positive rather than negative terms, and it is assumed here that this tendency can carry over to evaluations of (humanlike) non-human agents. Moreover, if a non-human agent is perceived as humanlike, it is likely that a human observer would perceive himself or herself as similar to the agent, and perceived similarity typically has a positive influence on evaluations (Cialdini, 2007). It is also possible that perceived humanness increases the extent to which we humans find it comfortable to rely on a non-human agent (Shi *et al.*, 2020), which in turn can boost evaluations of the agent.

In terms of findings in previous studies, a positive association between the extent to which a human person is imbued with humanness and evaluations of the person has been found by Kozak *et al.* (2006) and Söderlund (2020). Moreover, in consumer contexts, several studies have identified a positive influence of anthropomorphizing an offer on evaluations of the offer (e.g. Aggarwal and McGill, 2007; Delbaere *et al.*, 2011; Golossenko *et al.*, 2020). Hence the following is hypothesized for the situation in which a service robot is engaging in privacy-violating behavior by sharing personal information about a human:

- H3. Perceived robot humanness is positively associated with the overall evaluation of the robot.

The assumptions made so far for H1–H3 comprise a view of a service robot's privacy-violating behavior as indirectly influencing the overall evaluation of the robot. That is to say, given H1–H3, it is assumed that the effect of the robot's privacy-violating behavior on the overall evaluation of the robot is sequentially mediated by perceived morality and perceived humanness (see Figure 1). Other (and not hypothesized) mediators are indeed also possible, so in an attempt to examine if the sequence implied by H1–H3 survives an empirical test allowing also for other mediators (which would be indicated by a significant direct association between a service robot's privacy-violating behavior and the overall evaluation of the robot), the following is hypothesized:

- H4. The influence of robotic privacy-violating behavior on the evaluation of the robot is sequentially mediated by perceived morality and perceived humanness.

3. Research method

3.1 Research design, stimulus material and participants

A between-subjects experiment was employed to manipulate a service robot's sharing of personal information. The robot manipulations were made with a Wizard of Oz approach, which means that robot behavior was not autonomous, but it was made to appear as autonomous by the experimenter (cf. Broadbent, 2017; Riek, 2012). When a robot displays autonomous behavior, then, no human should appear to be "in the loop". This approach has been used in robot morality-related experiments by, for example, Briggs and Scheutz (2014), Hedaoo *et al.* (2019), Jackson *et al.* (2020) and Syrdal *et al.* (2007).

The basis for the manipulations was the creation of a situation involving a human-to-robot interaction. A script for such a situation was developed by the author. In this script, the interaction took place in an office environment comprising a male human employee who asked for help from a service robot (described as able to move around in the office and able to record, store and access peoples' conversations). In relation to a traditional service encounter (involving the interaction between a human frontline employee and a human consumer), the setting in the present study involves a backstage, internal interaction in which the service provided by a robot is intended for employees. Such employee-to-robot interactions have been identified as one of the main challenges for HR managers (Xu *et al.*, 2020). More specifically, the human employee asked the robot to disclose information about what the office manager had said about him, and about his current project, to other people. Thus the robot's response had to do with how it dealt with the manager's privacy. This setting was supposed to reflect that not discussing with others what one is told in confidence by a specific person is an important privacy-related social norm (Argyle *et al.*, 1985) and that listening activities is one route to privacy violation by robots (Rueben and Smart, 2016).

In the next step, videos were produced to depict the scripted situation. A Nao-looking robot was used in the role of the service robot, and it was given a synthetic voice by the means of a software that translates text to speech. The latter was inspired by the use of robot stimuli in Pan *et al.* (2015). Three versions of how the video ended were created. In the first version, to represent a relatively low level of privacy violation, the service robot did not disclose the requested information. In the second and the third versions, and to represent a relatively high level of privacy violation, the robot did disclose the requested information. The only difference between these two versions was that the second version comprised negatively valenced information, while the third version comprised positively valenced information. This aspect was added to control for the possibility that the valence of shared personal

information may (1) affect the extent to which sharing is seen as a norm violation and (2) be influential when it comes to overall evaluations. The latter possibility should be seen in the light of studies in a human-to-human context suggesting that the valence of a sender's information about other persons may influence evaluations of the sender in a valence-congruent way (Ruggiero *et al.*, 2020; Wyer *et al.*, 1990). The script versions and links to the three videos are provided in Experimental stimuli.

The participants were recruited from Prolific, an online panel built for research purposes (cf. Palan and Schitter, 2018). Qualtrics, a software for collecting data online, was used to randomly allocate the participants to one of the three video versions, and after watching they answered a set of questions comprising measures of the variables in the hypotheses. Three hundred and fifty-five participants completed the study. However, 2 participants were not able to watch the video, and 6 participants failed to respond correctly to an attention check item, and they were removed. Thus the analysis was based on those that remained ($n = 347$, $M_{age} = 34.26$; 209 women and 138 men).

3.2 Measurements

All items were scored on 10-point scales, and Cronbach's alpha (CA) was used to estimate the reliability of the multi-item measures. Discriminant validity was assessed with SmartPLS 4.0 and its module for the heterotrait-monotrait ratio of correlations. All ratios were <0.90 , thus indicating acceptable levels of discriminant validity.

As a manipulation check of the extent to which the videos were able to represent different levels of *privacy violation* (low vs high) by the robot, the participants were asked "What characterized the robot in the video?", followed by the items "It did not care about others' privacy-It cared about others' privacy", "Its attitude was that information wants to be free-Its attitude was that some things should not be shared", and "It acted as if no secrets should exist-It acted as if secrets should exist" (1 = Do not agree at all; 10 = Agree completely; CA = 0.94). Please notice that low scores reflect a view that privacy has been violated, while high scores reflect a view that privacy has been respected.

Perceived morality was measured with the items "The robot could understand negative and positive consequences of its behavior", "The robot could distinguish between what is good and what is bad", "The robot understood what could be harmful to people", and "The robot appeared to have had a sense of what is fair" (1 = Do not agree at all, 10 = Agree completely; CA = 0.96). Items of this type have been used by Gray *et al.* (2012).

Perceived humanness of the robot was measured with the items "The robot appeared very much as a human", "The robot was humanlike", and "The robot acted like humans typically do" (1 = Do not agree at all, 10 = Agree completely; CA = 0.91). Similar items have been used by, for example, Aggarwal and McGill (2007), Choi *et al.* (2019) and Söderlund (2020).

The participants' *overall evaluation of the robot* was measured with the question "What is your overall impression of the service robot in the video?" followed by the adjective pairs "bad-good", "dislike it-like it" and "unpleasant-pleasant" (CA = 0.94). Items of this type are commonly employed to capture consumers' overall evaluations of an object, in terms of the attitude towards the object, and they have been used by, for example, MacKenzie and Lutz (1989). As a validity check, and given that the attitude towards an object is assumed to influence behavioral intentions vis-à-vis the object, the participants were asked the following: "To what extent would you like to borrow the service robot in the video for a week, free of charge, to explore it in more detail in your home or work place?" (1 = Not at all, 10 = Very much). The attitude measure was positively and significantly associated with the responses to the intention item ($r = 0.43, p < 0.01$), which provides evidence for the predictive validity of the attitude measure.

Finally, to assess the realism of the service robot (which was portrayed as having more advanced capabilities than most existing robots for the purpose of the experiment), the participants were asked to respond to the following item: “Robots with social capabilities of the type displayed in the video . . .”. It was followed by the response alternatives “. . . will never exist” (chosen by 6 participants), “. . . exist already” (chosen by 228 participants) and “. . . will exist in the future” (chosen by 113 participants). This thus indicates that the robot appeared as realistic for the majority of the participants. The three video versions did not produce significantly different responses to the realism check item ($\chi^2 = 5.52, p = 0.24$).

4. Analysis and results

4.1 Manipulation check

A manipulation check with the privacy violation variable (please recall that it was scored so that a low score reflects a high level of perceived privacy violation) showed that it reached a higher level in the low privacy violation condition ($M = 8.09$) compared to the conditions representing a high level of privacy violation with a negative charge of the disclosed information ($M = 2.51$) and a high level of privacy violation with a positive charge of the disclosed information ($M = 2.81$). A one-way ANOVA indicated that all means were not equal ($F = 297.14, p < 0.01$). An assessment with Sheffé’s post hoc test indicated that the level of the privacy violation variable in the low privacy violation condition was significantly higher than in both the negatively charged ($p < 0.01$) and the positively charged ($p < 0.01$) high privacy violation conditions. The (low) levels of the privacy violation variable were not significantly different between the two high violation conditions ($p = 0.41$). This suggests that the manipulation of low vs high privacy violations worked as intended (and that it did not matter if the disclosed information had a negative or a positive charge).

The manipulation also produced different levels of the other variables in the hypotheses (i.e. morality, humanness and the overall evaluation) along the same lines: for each of these variables, the low privacy violation condition produced significantly higher levels than the two high privacy violation conditions (see Table 1). These outcomes, then, suggest that people are sensitive to when someone’s privacy is violated – even when the transgressor is a non-human agent.

4.2 Testing the hypotheses

The hypotheses were tested with a structural equation modeling approach (SmartPLS 4.0 was used). For these tests, the proposed model comprised the associations depicted in Figure 1 as well as a (non-hypothesized) direct association between privacy-violating behavior and the overall evaluation. This direct association was added to enable a test of the mediation hypothesis within the frame of the proposed model (more about this follows below). Privacy-violating behavior in the proposed model was scored as 1 = low level and 2 = high level as a reflection of the experimental conditions for the participants (i.e. the two high

Table 1.
Group means for the
three conditions

Response variable	Low privacy violation	High privacy violation (negative charge)	High privacy violation (positive charge)
Morality	6.54	2.49	2.85
Humanness	4.16	3.18	3.29
Overall evaluation	6.68	5.27	5.59

Source(s): Table created by author

privacy violation conditions were pooled, since they did not generate significant differences between any of the variables in the hypotheses). This proposed model explained 31% of the variation in the overall evaluation of the service robot. The model's fit was as follows: SRMR = 0.04 and NFI = 0.92.

For H1–H3, the path coefficients are reported in Table 2. As hypothesized, the privacy violation–morality association was negative, while the morality–humanness association and the humanness–overall association were positive. Each association was significant, thus providing support for H1–H3.

With respect to the mediation hypothesis (H4), stating that the influence of privacy-violating behavior by a service robot on the overall evaluation of the robot is sequentially mediated by perceived morality and perceived humanness, the SEM-based procedure suggested by Nitzl *et al.* (2016) and Sarstedt *et al.* (2020) was followed. More specifically, and as already indicated, a direct association between privacy-violating behavior and the overall evaluation of the robot was included in the proposed model. This was done to be able to control for a direct effect and, if the indirect link is significant, to assess the type of mediation. With this approach, mediation is at hand if there is a significant indirect effect between an independent variable and a dependent variable, which is indicated by the confidence interval for the coefficient for the indirect effect (it should not comprise a zero). Nitzl *et al.* (2016) recommends a biased-corrected confidence interval for the assessment, and this was used in the present mediation analysis. This analysis showed that there was a significant indirect effect ($b = -0.34$, $t = 6.89$; $p < 0.01$) of privacy-violating behavior on the overall evaluation of the robot via perceived morality and perceived humanness. This means that H4 was supported. It should be noted that the direct privacy violation–overall evaluation association was significant, too ($b = -0.44$, $t = 4.73$, $p < 0.01$). This, then, indicates complementary mediation (cf. Zhao *et al.*, 2010) and thus that other (and not hypothesized) variables are likely to serve as additional mediators.

5. Discussion

5.1 Contributions

As noted by several service scholars, the service encounter is fundamentally changing due to rapid evolutions in technology, and this calls for updated perspectives (Larivière *et al.*, 2017; Wirtz *et al.*, 2021). The present paper contributes to such perspectives by examining one type of the “service encounter 2.0” (Larivière *et al.*, 2017) in which the human user is interacting with a service robot. More specifically, in relation to many researchers’ optimistic view of service robots, such as Wirtz *et al.* (2021), who assume that service robots will bring unprecedented improvements to the customer experience, the present study indicates that not everything that robots do would have a positive charge. The present study also contributes to the area of adoption of new services. Several existing studies in this field have identified privacy concerns as a factor that can mitigate adoption of smart services (e.g. Chouk and Mani, 2019; Hong *et al.*, 2020) and attenuate the acceptance of technologies to be used in service encounters (Pizzi and Scarpi, 2020), and to this the present study adds an examination of what happens when a robotic privacy violation takes place (and mechanisms

Hypothesis	<i>b</i>	<i>t</i>	<i>p</i>
H1: Privacy violation – Morality	–1.41	14.38	<0.01
H2: Morality – Humanness	0.50	10.82	<0.01
H3: Humanness – Evaluation	0.48	11.38	<0.01

Source(s): Table created by author

Table 2.
Path coefficients for the
H1–H3 associations

in terms of perceived morality and humanness). In relation to existing studies of service robots in the field of service research, it should be noted that most studies comprise users who are consumers. That is to say, despite the fact that service robots are expected to become more common as co-workers (Le *et al.*, 2022), and thus they will be providing service to employees, relatively few studies have examined employee-to-robot service encounters.

The present study should also be seen as a response to calls for more examinations of robot morality – sometimes also referred as machine ethics, machine morality and artificial morality – and to specific calls for privacy-related issues in human-to-robot interactions (Calo, 2012; Moor, 2006; Rueben and Smart, 2016). Several authors have argued that robots are capable of violating human privacy in many ways (e.g. Rueben *et al.*, 2018) and the present study shows that human participants are indeed able to perceive a robot's sharing of personal information about a human in privacy-violation terms. Given that privacy violations are norm transgressions, much existing research shows that humans who break norms are subject to immediate negative reactions and are likely to be condemned in moral terms (e.g. Haidt, 2001; Malle, 2014). The present study indicates a similar reaction pattern when the transgressor is a robot, in the sense that robot-generated privacy violations attenuated perceived robot morality. It should be underlined that one of the conditions in the present study involved the preservation of human privacy. That is to say, in this condition, the privacy-friendly robot refused to obey the request of the human user, and by doing this the robot was able to boost perceptions about its respect for human privacy (and, as consequence, perceived morality was boosted). These results should be seen in the light of Bigman and Gray (2018) who found that people in general seem to be aversive to allowing machines making morally-relevant decisions (when these decisions have to do with issues of life and death). Privacy issues have less serious consequences than life and death issues, so the part of the results of the present study showing a positive impact of norm compliance could indicate that people may not think that all morally-relevant decisions should be reserved for humans only. It should also be noted that the perceived level of robot morality in the low privacy violation group (i.e. $M = 6.54$) appears to be higher than in some previous studies (e.g. Weisman *et al.*, 2017). The present study, then, offers evidence that service robots can be perceived as possessing more than a very low level of morality. In any event, and with respect to perceptions of privacy violations and their impact on overall evaluations, the present study adds further empirical evidence in support of the general paradigm assuming that we humans tend to react to (humanlike) non-humans in ways that are similar to how we react to real humans (Epley, 2018; Reeves and Nass, 1996).

Furthermore, with respect to perceived morality and its influence on perceived humanness, the present study produced a similar pattern as studies conducted in a human-to-human context (e.g. Bastian *et al.*, 2013). That is to say, a perceived lack of morality in humans typically goes hand in hand with dehumanization (and reduced levels of overall evaluations). The same pattern – a positive association between perceived morality and perceived humanness – has also been obtained for (non-embodied) virtual agents used in consumer settings (Söderlund and Oikarinen, 2021). That is to say, the present study shares the assumption with other studies that perceived morality influences perceived humanness.

It should be underlined, however, that some other studies view the link between morality and perceived humanness of non-human agents in terms of reversed causality; it has been assumed that the more humanlike a non-human agent appears, the more it is ascribed mind (e.g. Bigman *et al.*, 2019). These two views of the time asymmetry are not necessarily contradictory. Presumably, humanlike features (e.g. a head, arms, legs and a voice) provide the observer with an initial suggestion of humanness, which in turn makes it easier to see a mind (including the capability to reason in moral terms) and, in the next step, the perceived presence of morality boosts perceived humanness further. Yet, these are speculations. To come to terms with the identity of a mediation variable as a cause or an effect vis-à-vis

another mediation variable, one would need to move beyond an approach in which mediators are measured (as in the present study) and use a research design in which mediators are manipulated by the experimenter (Spencer *et al.*, 2005).

As for consequences of humanness perceptions of non-human agents, previous research has stressed that the extent to which a robot is humanlike has implications for human-to-robot interactions (e.g. Broadbent, 2017; Murphy *et al.*, 2019). Indeed, the existing literature has identified several consequences of perceived humanness (and/or anthropomorphism) of artificial agents. The present study adds further evidence that overall evaluations is one of these consequences. This variable, however, is more common outside the robot literature than inside it. Instead, existing studies on the effects of humans' perceptions of robot morality have included, for example, response variables such as trust, perceived intelligence, comfort and social attraction (Banks, 2019; Malle *et al.*, 2016). The relatively low level of interest in overall evaluations in the field of robotics is unfortunate, because an evaluation is a central variable in many theoretical fields in which various aspects of human information processing are examined (e.g. social psychology and marketing). Using overall evaluation variables also in research on human reactions to robots, then, would facilitate contact between theories of robot behavior and theories of human behavior and thereby enable the development of richer theories for human-to-robot interactions.

5.2 Implications for decision-makers

It should be recalled that the experimental setting involved a workplace context with an employee-to-robot interaction (i.e. the robot provided service to an employee). Given this setting, the findings have implications for two types of decision makers. The first are those who make decisions about service robots' abilities and characteristics in firms producing robots; the second are those who make decisions about robots (e.g. HR managers; cf. Xu *et al.*, 2020) in firms that adopt robots as a means to provide service to employees. Given that overall evaluations of a service robot have behavioral implications for existing and potential users, these decision-makers should pay attention to how robots deal with issues that are likely to be perceived as privacy violations by users. A main challenge, however, is to create a portfolio of privacy-preserving norms that a robot should follow (Vanderelst and Willems, 2019). This is challenging, because many such norms are subject to disagreement within and between legal academics, philosophers, morality scholars and the general public. Additional complexity is added by the fact that the right to privacy is limited by other rights in many situations (Sorell and Draper, 2017). Another challenge is to program a robot so that it does not violate the selected norms (cf. Vanderelst and Willems, 2019; Wallach, 2010).

Nevertheless, given that a set of norms can be identified, a general approach to avoid violations is so-called privacy by design, which has to do with embedding privacy as a default setting into a system (Ishii, 2019). Several authors have suggested specific ways to restrict the potential of robots to violate human privacy along such lines, and they may serve as a point of departure for at least the first steps for those whose task it is to create privacy-preserving robots. For example, human privacy can be preserved in technical ways, such as imposing limits to where robots can go (e.g. bedrooms) and to what their cameras are allowed to see (Rueben *et al.*, 2018). Robots can also be programmed to signal that they are in a recording mode and to refuse to disclose personal information about humans.

It should also be noted that the initiative for privacy violation in the present study did not come from the stimulus robot – it was a curious human who set the process in motion. Given robots that move around in environments with humans, collect data from them, store these data and can retrieve them later on, one may foresee an almost endless number of situations in which humans would be very interested in getting access to such data. Serious attempts to come to terms with privacy issues in relation to robots should therefore target also human

users with respect to codes of conduct for what it is legitimate to demand from a robot. One possibility is to make robots themselves involved in this task; for example, robots could be programmed to nudge a human user in the direction of not demanding immoral actions (Borenstein and Arkin, 2016).

However, an additional aspect may complicate matters for decision makers who want to create privacy-preserving robots: we humans typically have a powerful motivation to disclose information about ourselves (Acquisti *et al.*, 2015). This motivation is likely to be particularly salient in the type of office environment that was used as a setting in the present study. That is to say, service robots that share the environment with humans and collect data about them, may serve as an important and welcomed communication instrument. An overly cautious approach to privacy, then, may reduce the potential for service robots to become a useful means for those who actually want to be noticed and listened to.

5.3 Limitations and suggestions for further research

The present study examined one of several ways in which a robot can violate human privacy (i.e. the robot disclosed information about a human to another human). Further studies, however, are needed to examine other privacy violation behaviors, for example, how robots collect and store data – as a well as other aspects of access to these data than disclosure by the robot itself (Rueben and Smart, 2016). Moreover, it should again be underlined that the present study comprised a curious human user who set the privacy-violation process in motion. Such violations, however, can also be initiated by a robot. For example, they can be initiated by a proactive robot in order to provide the user with (what the robot thinks) is valuable information, and they can occur as errors. Such circumstances may not produce the same reactions as in the present study; thus, they need attention in further research. In addition, the present study comprised only one way for a robot to refuse to engage in privacy violations. Such refusals, however, can be made in several ways (e.g. with or without motivations and with different levels of politeness), and more research is needed to identify if different types of refusals produce different effects. Research along the lines just mentioned may, as a by-product, also contribute to our knowledge of privacy violations in a human-to-human context. Although attempts have been made to create taxonomies of privacy violations in that context (e.g. Leino-Kilpi *et al.*, 2001), the existing literature does not yet comprise a detailed catalogue of how serious they are perceived to be and what responses they evoke by those who are victims (or by observers). In other words, examinations of robot morality can increase our understanding of human morality (Misselhorn, 2018; Moor, 2006; Wallach, 2010).

Moreover, many service robot attributes can influence overall evaluations. Several studies have examined such attributes in terms of various versions of technology acceptance models. Examples of antecedent variables are the extent to which a new technology is perceived to be useful and is perceived to be easy to use (Davis, 1993), and they have indeed been shown to contribute to the variation in dependent variables also in the case of service robots (e.g. Park *et al.*, 2021; Zhong *et al.*, 2020). The present study, however, did not include variables of this type. One may perhaps question the value of once again confirming that such variables contribute to robot-related outcome variables, but they can serve the purpose of being control variables in tests of variables of the type used in the present study. That is to say, a limitation of the present study is that it was not able to show that H1–H3 survived a test in which also well-established antecedents to overall evaluations of a technology were present.

As for mechanisms behind the influence of privacy violations on overall evaluations, the present study indicates that perceived morality and perceived humanness were mediators. Other mediators, however, are likely to exist, too. In the present study, morality was seen as a high-level capacity that makes moral sense-making at lower levels possible. An example

of the latter is moral judgment (i.e. the assessment of the extent to which one particular act is wrong or right). Another example is understanding that several norms can exist for one particular act and that they may be in conflict. Thus, several morality-related variables may be elicited by a norm-violating act, and thereby they are additional candidates for examining reactions to robot behavior with moral implications. Moreover, [Gray et al. \(2014\)](#) argue that a norm violation is likely to be followed by – or to be intertwined with – not only moral judgement but also by intuitions about a harmed victim. This means that the identity and perceptions of the victim of privacy violations may influence morality assessments and downstream variables such as the attitude towards the transgressor. In the manipulations in the present study, the victim was an office manager whose thoughts about another person were shared (or not shared) by the stimulus robot, and one may perhaps assume that the harm for this victim in the privacy violation conditions was not extensive. After all, most managers are likely to think that employees pay attention to what they say about other employees and to think that employees are likely to pass this on and discuss it further. Other types of victims, however, may be perceived as being more harmed or may be cared for more by observers of privacy violations, which in turn may influence attributions of morality to the transgressor. Moreover, attributions of morality are likely to be influenced by other aspects of having mind, particularly agency and emotionality ([Gray et al., 2012, 2014](#)), so perceptions of such additional mind aspects are needed to enrich our knowledge about the process by which robots are attributed morality.

It should be underlined that privacy violations can involve several norms. In the present study, for example, and when the robot refused to obey the human user, such a refusal is itself a morality-charged decision, because it may violate norms about how robots should behave in relation to the user. Indeed, the refusal in the present study may violate one of the seminal robot laws launched by science-fiction writer Isaac Asimov, which has been discussed in the robotics literature: a robot must obey the orders given to it by human beings ([Ishii, 2019](#)). In any event, when a robot has to navigate in a landscape comprising several norms, more precision in capturing humans' responses to robots may be obtained by the use of explicit measures of perceptions of violations/compliance in relation to each norm that is involved.

Potential moderators (i.e. variables that may attenuate or boost the impact of robots' privacy violation behaviors on overall evaluations) should also be examined in further studies. First, it should be noted that the participants in the present study had a relatively detached role in relation to what happened in the videos (i.e. it was not the *participants'* privacy that was violated or preserved). A study in which the participants' own privacy is at stake may therefore produce stronger effects than those that were obtained in the present study. Thus one's role as a mere observer or as an actual victim may moderate the impact of privacy violations. Second, in the present study the information that was disclosed was a person's thoughts and evaluations of another person. This is sometimes referred to as psychological privacy as opposed to several other types of privacy with respect to what information that is in focus ([Leino-Kilpi et al., 2001](#)). Similarly, the personal information that is disclosed to others can vary in terms of how personal it is ([Derlega and Chaikin, 1977](#)); it can range from being superficial (e.g. one's favorite food) to highly personal and intimate matters (e.g. personal inadequacies and one's sex life). Thus, different types of personal information can be involved in privacy violations, and the effect they produce on downstream variables is unlikely to be the same. Third, the extent to which one particular act with privacy-violation potential is seen as a transgression can be subject to individual variation ([Acquisti et al., 2015](#)). For example, individuals' moral convictions regarding an act (the extent to which it is connected to their fundamental beliefs about right and wrong; [Skitka, 2010](#)) can vary, and such convictions are likely to affect the impact of privacy violations on various response variables. Finally, (a high level of) privacy has been associated with many benefits. But it can also produce harm. For example, it can support illegitimate goals ([Margulis, 2003](#)), such

as when a future tyrant or terrorist is provided room for developing thoughts and plans about how others should be oppressed or destroyed. Thus, there are situations in which an individual's loss of control of information about him/herself can be highly valuable for others, and when robots break norms in such situations, it may boost how humans evaluate them.

References

- Acquisti, A., Brandimarte, L. and Loewenstein, G. (2015), "Privacy and human behavior in the age of information", *Science*, Vol. 347 No. 6221, pp. 509-514.
- Aggarwal, P. and McGill, A.L. (2007), "Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products", *Journal of Consumer Research*, Vol. 34 No. 4, pp. 468-479.
- Allen, C., Wallach, W. and Smit, I. (2006), "Why machine ethics?", *IEEE Intelligent Systems*, Vol. 21 No. 4, pp. 12-17.
- Argyle, M., Henderson, M. and Furnham, A. (1985), "The rules of social relationships", *British Journal of Social Psychology*, Vol. 24 No. 2, pp. 125-139.
- Astor, M. (2017), "Your Roomba may be mapping your home, collecting data that could be shared", *The New York Times*.
- Banks, J. (2019), "A perceived moral agency scale: development and validation of a metric for humans and social machines", *Computers in Human Behavior*, Vol. 90 January, pp. 363-371.
- Baruh, L., Secinti, E. and Cemalcilar, Z. (2017), "Online privacy concerns and privacy management: a meta-analytical review", *Journal of Communication*, Vol. 67 No. 1, pp. 26-53.
- Bastian, B., Loughnan, S., Haslam, N. and Radke, H.R. (2012), "Don't mind meat? The denial of mind to animals used for human consumption", *Personality and Social Psychology Bulletin*, Vol. 38 No. 2, pp. 247-256.
- Bastian, B., Denson, T.F. and Haslam, N. (2013), "The roles of dehumanization and moral outrage in retributive justice", *PloS One*, Vol. 8 No. 4, e61842.
- Baumeister, R.F., Bratslavsky, E., Finkenauer, C. and Vohs, K.D. (2001), "Bad is stronger than good", *Review of General Psychology*, Vol. 5 No. 4, pp. 323-370.
- Belanche, D., Casaló, L.V., Flavián, C. and Schepers, J. (2020), "Robots or frontline employees? Exploring customers' attributions of responsibility and stability after service failure or success", *Journal of Service Management*, Vol. 31 No. 2, pp. 267-289.
- Bigman, Y.E. and Gray, K. (2018), "People are averse to machines making moral decisions", *Cognition*, Vol. 181 December, pp. 21-34.
- Bigman, Y.E., Waytz, A., Alterovitz, R. and Gray, K. (2019), "Holding robots responsible: the elements of machine morality", *Trends in Cognitive Sciences*, Vol. 23 No. 5, pp. 365-368.
- Bitner, M.J., Booms, B.H. and Tetreault, M.S. (1990), "The service encounter: diagnosing favorable and unfavorable incidents", *Journal of Marketing*, Vol. 54 January, pp. 71-84.
- Borenstein, J. and Arkin, R. (2016), "Robotic nudges: the ethics of engineering a more socially just human being", *Science and Engineering Ethics*, Vol. 22 No. 1, pp. 31-46.
- Briggs, G. and Scheutz, M. (2014), "How robots can affect human behavior: investigating the effects of robotic displays of protest and distress", *International Journal of Social Robotics*, Vol. 6 No. 3, pp. 343-355.
- Broadbent, E. (2017), "Interactions with robots: the truths we reveal about ourselves", *Annual Review of Psychology*, Vol. 68, pp. 627-652.
- Calo, M.R. (2012), "Robots and privacy", in Lin, P., Abney, K. and Bekey, G.A. (Eds.), *Robot Ethics-The Ethical and Social Implications of Robotics*, MIT Press, Cambridge, MA, pp. 187-201.
- Choi, S., Liu, S.Q. and Mattila, A.S. (2019), "How may I help you? Says a robot: examining language styles in the service encounter", *International Journal of Hospitality Management*, Vol. 82 September, pp. 32-38.

- Chouk, I. and Mani, Z. (2019), "Factors for and against resistance to smart services: role of consumer lifestyle and ecosystem related variables", *Journal of Services Marketing*, Vol. 33 No. 4, pp. 449-462.
- Cialdini, R.B. (2007), *Influence: The Psychology of Persuasion*, Collins, New York.
- Coeckelbergh, M. (2010), "Moral appearances: emotions, robots, and human morality", *Ethics and Information Technology*, Vol. 12 No. 3, pp. 235-241.
- Crnkovic, G.D. and Çürüklü, B. (2012), "Robots: ethical by design", *Ethics and Information Technology*, Vol. 14 No. 1, pp. 61-71.
- Damiano, L. and Dumouchel, P. (2018), "Anthropomorphism in human-robot co-evolution", *Frontiers in Psychology*, Vol. 9 March, 468.
- Davis, F.D. (1993), "User acceptance of information technology: system characteristics, user perceptions and behavioral impacts", *International Journal of Man-Machine Studies*, Vol. 38 No. 3, pp. 475-487.
- Delbaere, M., McQuarrie, E.F. and Phillips, B.J. (2011), "Personification in advertising", *Journal of Advertising*, Vol. 40 No. 1, pp. 121-130.
- Derlega, V.J. and Chaikin, A.L. (1977), "Privacy and self-disclosure in social relationships", *Journal of Social Issues*, Vol. 33 No. 3, pp. 102-115.
- Epley, N. (2018), "A mind like mine: the exceptionally ordinary underpinnings of anthropomorphism", *Journal of the Association for Consumer Research*, Vol. 3 No. 4, pp. 591-598.
- Epley, N., Waytz, A., Akalis, S. and Cacioppo, J.T. (2008), "When we need a human: motivational determinants of anthropomorphism", *Social Cognition*, Vol. 26 No. 2, pp. 143-155.
- Epley, N., Schroeder, J. and Waytz, A. (2013), "Motivated mind perception: treating pets as people and people as animals", in Gervais, S. (Ed.), *Objectification and (de)humanization*, Springer, New York.
- Fowler, G. (2019), "What does your car know about you? We hacked a Chevy to find out", *The Washington Post*.
- Golossenko, A., Pillai, K.G. and Aroean, L. (2020), "Seeing brands as humans: development and validation of a brand anthropomorphism scale", *International Journal of Research in Marketing*, Vol. 37 No. 4, pp. 737-755.
- Goodwin, G.P., Piazza, J. and Rozin, P. (2014), "Moral character predominates in person perception and evaluation", *Journal of Personality and Social Psychology*, Vol. 106 No. 1, pp. 148-168.
- Gray, K., Young, L. and Waytz, A. (2012), "Mind perception is the essence of morality", *Psychological Inquiry*, Vol. 23 No. 2, pp. 101-124.
- Gray, K., Schein, C. and Ward, A.F. (2014), "The myth of harmless wrongs in moral cognition: automatic dyadic completion from sin to suffering", *Journal of Experimental Psychology: General*, Vol. 143 No. 4, pp. 1600-1615.
- Guglielmo, S. (2015), "Moral judgment as information processing: an integrative review", *Frontiers in Psychology*, Vol. 6, doi: [10.3389/fpsyg.2015.01637](https://doi.org/10.3389/fpsyg.2015.01637).
- Haidt, J. (2001), "The emotional dog and its rational tail: a social intuitionist approach to moral judgment", *Psychological Review*, Vol. 108 No. 4, pp. 814-834.
- Harenski, C.L. and Kiehl, K.A. (2011), "Emotion and morality in psychopathy and paraphilias", *Emotion Review*, Vol. 3 No. 3, pp. 299-301.
- Haslam, N. (2006), "Dehumanization: an integrative review", *Personality and Social Psychology Review*, Vol. 10 No. 3, pp. 252-264.
- Hedao, S., Williams, A., Wadgaonkar, C. and Knight, H. (2019), "A robot barista comments on its clients: social attitudes toward robot data use", *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, pp. 66-74.
- Hong, A., Nam, C. and Kim, S. (2020), "What will be the possible barriers to consumers' adoption of smart home services?", *Telecommunications Policy*, Vol. 44 No. 2, doi: [10.1016/j.telpol.2019.101867](https://doi.org/10.1016/j.telpol.2019.101867).

- Ishii, K. (2019), "Comparative legal study on privacy and personal data protection for robots equipped with artificial intelligence: looking at functional and technological aspects", *AI and Society*, Vol. 34 No. 3, pp. 509-533.
- Jackson, R.B., Williams, T. and Smith, N. (2020), "Exploring the role of gender in perceptions of robotic noncompliance", *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 559-567.
- Kahn, P.H. Jr, Ishiguro, H., Friedman, B., Kanda, T., Freier, N.G., Severson, R.L. and Miller, J. (2007), "What is a human? Toward psychological benchmarks in the field of human-robot interaction", *Interaction Studies*, Vol. 8 No. 3, pp. 363-390.
- Kim, Y. and Sundar, S.S. (2012), "Anthropomorphism of computers: is it mindful or mindless?", *Computers in Human Behavior*, Vol. 28 No. 1, pp. 241-250.
- Kirmani, A. (1997), "Advertising repetition as a signal of quality: if it's advertised so much, something must be wrong", *Journal of Advertising*, Vol. 26 No. 3, pp. 77-86.
- Kozak, M.N., Marsh, A.A. and Wegner, D.M. (2006), "What do I think you're doing? Action identification and mind attribution", *Journal of Personality and Social Psychology*, Vol. 90 No. 4, pp. 543-555.
- Krupp, M.M., Rueben, M., Grimm, C.M. and Smart, W.D. (2017), "A focus group study of privacy concerns about telepresence robots", *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, IEEE, pp. 1451-1458.
- Larivière, B., Bowen, D., Andreassen, T.W., Kunz, W., Sirianni, N.J., Voss, C. and De Keyser, A. (2017), "Service Encounter 2.0": an investigation into the roles of technology, employees and customers", *Journal of Business Research*, Vol. 79 October, pp. 238-246.
- Le, K.B.Q., Sajtos, L. and Fernandez, K.V. (2022), "Employee-(ro)bot collaboration in service: an interdependence perspective", *Journal of Service Management*, Vol. 34 No. 2, pp. 176-207, doi: [10.1108/JOSM-06-2021-0232](https://doi.org/10.1108/JOSM-06-2021-0232).
- Lee, M.K., Tang, K.P., Forlizzi, J. and Kiesler, S. (2011), "Understanding users' perception of privacy in human-robot interaction", *2011 6th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, pp. 181-182.
- Leino-Kilpi, H., Välimäki, M., Dassen, T., Gasull, M., Lemonidou, C., Scott, A. and Arndt, M. (2001), "Privacy: a review of the literature", *International Journal of Nursing Studies*, Vol. 38 No. 6, pp. 663-671.
- Lu, V.N., Wirtz, J., Kunz, W.H., Paluch, S., Gruber, T., Martins, A. and Patterson, P.G. (2020), "Service robots, customers and service employees: what can we learn from the academic literature and where are the gaps?", *Journal of Service Theory and Practice*, Vol. 30 No. 3, pp. 361-391.
- Lynskey, D. (2019), "Alexa, are you invading my privacy? – The dark side of our voice assistants", *The Guardian*.
- MacKenzie, S.B. and Lutz, R.J. (1989), "An empirical examination of the structural antecedents of attitude toward the ad in an advertising pretesting context", *Journal of Marketing*, Vol. 53 No. 2, pp. 48-65.
- Magi, T.J. (2011), "Fourteen reasons privacy matters: a multidisciplinary review of scholarly literature", *The Library Quarterly*, Vol. 81 No. 2, pp. 187-209.
- Malle, B.F. (2014), "Moral competence in robots?" in Seibt, J., Hakli, R. and Nørskov, M. (Eds), *Sociable Robots and the Future of Social Relations: Proceedings of Robo-Philosophy 2014*, pp. 189-198, IOS Press, Amsterdam.
- Malle, B.F., Guglielmo, S. and Monroe, A.E. (2014), "A theory of blame", *Psychological Inquiry*, Vol. 25 No. 2, pp. 147-186.
- Malle, B.F., Scheutz, M., Forlizzi, J. and Voiklis, J. (2016), "Which robot am I thinking about? The impact of action and appearance on people's evaluations of a moral robot", *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE.

- Margulis, S.T. (2003), "Privacy as a social issue and behavioral concept", *Journal of Social Issues*, Vol. 59 No. 2, pp. 243-261.
- Mikhail, J. (2007), "Universal moral grammar: theory, evidence and the future", *Trends in Cognitive Sciences*, Vol. 11 No. 4, pp. 143-152.
- Misselhorn, C. (2018), "Artificial morality. Concepts, issues and challenges", *Society*, Vol. 55 No. 2, pp. 161-169.
- Moor, J.H. (2006), "The nature, importance, and difficulty of machine ethics", *IEEE Intelligent Systems*, Vol. 21 No. 4, pp. 18-21.
- Moore, A.D. (2003), "Privacy: its meaning and value", *American Philosophical Quarterly*, Vol. 40 No. 3, pp. 215-227.
- Moorthy, S. and Zhao, H. (2000), "Advertising spending and perceived quality", *Marketing Letters*, Vol. 11 No. 3, pp. 221-233.
- Murphy, J., Gretzel, U. and Pesonen, J. (2019), "Marketing robot services in hospitality and tourism: the role of anthropomorphism", *Journal of Travel and Tourism Marketing*, Vol. 36 No. 7, pp. 784-795.
- Nissenbaum, H. (2004), "Privacy as contextual integrity", *Washington Law Review*, Vol. 79 No. 1, pp. 119-157.
- Nitzl, C., Roldan, J.L. and Cepeda, G. (2016), "Mediation analysis in partial least squares path modeling", *Industrial Management and Data Systems*, Vol. 116 No. 9, pp. 1849-1864.
- Nowak, K.L. and Biocca, F. (2003), "The effect of the agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments", *Presence: Teleoperators and Virtual Environments*, Vol. 12 No. 5, pp. 481-494.
- Palan, S. and Schitter, C. (2018), "Prolific.ac: a subject pool for online experiments", *Journal of Behavioral and Experimental Finance*, Vol. 17 March, pp. 22-27.
- Pan, Y., Okada, H., Uchiyama, T. and Suzuki, K. (2015), "On the reaction to robot's speech in a hotel public space", *International Journal of Social Robotics*, Vol. 7 No. 5, pp. 911-920.
- Park, S.S., Tung, C.D. and Lee, H. (2021), "The adoption of AI service robots: a comparison between credence and experience service settings", *Psychology and Marketing*, Vol. 38 No. 4, pp. 691-703.
- Pizzi, G. and Scarpi, D. (2020), "Privacy threats with retail technologies: a consumer perspective", *Journal of Retailing and Consumer Services*, Vol. 56 September, doi: [10.1016/j.jretconser.2020.102160](https://doi.org/10.1016/j.jretconser.2020.102160).
- Powers, A. and Kiesler, S. (2006), "The advisor robot: tracing people's mental model from a robot's physical attributes", *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp. 218-225.
- Reeves, B. and Nass, C. (1996), *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*, Cambridge University Press, New York.
- Riek, L.D. (2012), "Wizard of Oz studies in HRI: a systematic review and new reporting guidelines", *Journal of Human-Robot Interaction*, Vol. 1 No. 1, pp. 119-136.
- Rueben, M. and Smart, W.D. (2016), "Privacy in human-robot interaction: survey and future work", in *We Robot 2016: The Fifth Annual Conf. On Legal and Policy Issues Relating to Robotics*, University of Miami School of Law.
- Rueben, M., Aroyo, A.M., Lutz, C., Schmözl, J., Van Cleynenbreugel, P., Corti Agrawal, S. and Smart, W.D. (2018), "Themes and research directions in privacy-sensitive robotics", *2018 IEEE workshop on advanced robotics and its social impacts (ARSO)*, pp. 77-84, IEEE.
- Ruggiero, A., Parolin, E. and Ma, L. (2020), "The impact of gossip valence on children's attitudes towards gossipers", *Infant and Child Development*, Vol. 29 No. 4, e2180.
- Sarstedt, M., Hair, J.F. Jr, Nitzl, C., Ringle, C.M. and Howard, M.C. (2020), "Beyond a tandem analysis of SEM and PROCESS: use of PLS-SEM for mediation analyses", *International Journal of Market Research*, Vol. 62 No. 3, pp. 288-299.

- Scheutz, M. (2017), "The case for explicit ethical agents", *Ai Magazine*, Vol. 38 No. 4, pp. 57-64.
- Sears, D.O. (1983), "The person-positivity bias", *Journal of Personality and Social Psychology*, Vol. 44 No. 2, pp. 233-250.
- Shi, S., Gong, Y. and Gursay, D. (2020), "Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: a heuristic-systematic model", *Journal of Travel Research*, Vol. 60 No. 8, pp. 1714-1734, doi: [10.1177/0047287520966395](https://doi.org/10.1177/0047287520966395).
- Skitka, L.J. (2010), "The psychology of moral conviction", *Social and Personality Psychology Compass*, Vol. 4 No. 4, pp. 267-281.
- Smith, H.J., Dinev, T. and Xu, H. (2011), "Information privacy research: an interdisciplinary review", *MIS Quarterly*, Vol. 35 No. 4, pp. 989-1015.
- Söderlund, M. (2020), "Employee norm-violations in the service encounter during the corona pandemic and their impact on customer satisfaction", *Journal of Retailing and Consumer Services*, Vol. 57 November, doi: [10.1016/j.jretconser.2020.102209](https://doi.org/10.1016/j.jretconser.2020.102209).
- Söderlund, M. and Oikarinen, E.L. (2021), "Service encounters with virtual agents: an examination of perceived humanness as a source of customer satisfaction", *European Journal of Marketing*, Vol. 55 No. 13, pp. 94-121.
- Sorell, T. and Draper, H. (2017), "Second thoughts about privacy, safety and deception", *Connection Science*, Vol. 29 No. 3, pp. 217-222.
- Spencer, S.J., Zanna, M.P. and Fong, G.T. (2005), "Establishing a causal chain: why experiments are often more effective than mediational analyses in examining psychological processes", *Journal of Personality and Social Psychology*, Vol. 89 No. 6, pp. 845-851.
- Sullins, J.P. (2006), "When is a robot a moral agent?", *International Review of Information Ethics*, Vol. 6 No. 12, pp. 23-30.
- Sullins, J.P. (2011), "Introduction: open questions in roboethics", *Philosophy and Technology*, Vol. 24 No. 3, pp. 233-238.
- Swiderska, A. and Küster, D. (2020), "Robots as malevolent moral agents: harmful behavior results in dehumanization, not anthropomorphism", *Cognitive Science*, Vol. 44 No. 7, e12872.
- Syrdal, D.S., Walters, M.L., Otero, N., Koay, K.L. and Dautenhahn, K. (2007), "He knows when you are sleeping-privacy and the personal robot companion", *Proc. Workshop Human Implications of Human-Robot Interaction, Association for the Advancement of Artificial Intelligence (Aaai'07)*, pp. 28-33.
- Van Berkum, J.J., Holleman, B., Nieuwland, M., Otten, M. and Murre, J. (2009), "Right or wrong? The brain's fast response to morally objectionable statements", *Psychological Science*, Vol. 20 No. 9, pp. 1092-1099.
- van Leeuwen, F., Park, J.H. and Penton-Voak, I.S. (2012), "Another fundamental social category? Spontaneous categorization of people who uphold or violate moral norms", *Journal of Experimental Social Psychology*, Vol. 48 No. 6, pp. 1385-1388.
- van Kleef, G.A., Wanders, F., Stamkou, E. and Homan, A.C. (2015), "The social dynamics of breaking the rules: antecedents and consequences of norm-violating behavior", *Current Opinion in Psychology*, Vol. 6, pp. 25-31.
- Vanderelst, D. and Willems, J. (2019), "Can we agree on what robots should be allowed to do? An exercise in rule selection for ethical care robots", *International Journal of Social Robotics*, No. 12, pp. 1093-1102.
- Wallach, W. (2010), "Robot minds and human ethics: the need for a comprehensive model of moral decision making", *Ethics and Information Technology*, Vol. 12 No. 3, pp. 243-250.
- Weisman, K., Dweck, C.S. and Markman, E.M. (2017), "Rethinking people's conceptions of mental life", *Proceedings of the National Academy of Sciences*, Vol. 114 43, pp. 11374-11379.

- White, R.W. (1959), "Motivation reconsidered: the concept of competence", *Psychological Review*, Vol. 66 No. 5, pp. 297-333.
- Wirtz, J., Kunz, W. and Paluch, S. (2021), "The service revolution, intelligent automation and service robots", *The European Business Review*, Vol. 29 No. 5, pp. 38-44.
- Wyer, R.S., Budesheim, T.L. and Lambert, A.J. (1990), "Cognitive representation of conversations about persons", *Journal of Personality and Social Psychology*, Vol. 58 No. 2, pp. 218-238.
- Xu, S., Stienmetz, J. and Ashton, M. (2020), "How will service robots redefine leadership in hotel management? A Delphi approach", *International Journal of Contemporary Hospitality Management*, Vol. 32 No. 6, pp. 2217-2237.
- Zhao, X., Lynch, J.G. Jr and Chen, Q. (2010), "Reconsidering Baron and Kenny: myths and truths about mediation analysis", *Journal of Consumer Research*, Vol. 37 No. 2, pp. 197-206.
- Zhong, L., Zhang, X., Rong, J., Chan, H.K., Xiao, J. and Kong, H. (2020), "Construction and empirical research on acceptance model of service robots applied in hotel industry", *Industrial Management and Data Systems*, Vol. 212 No. 6, pp. 1325-1352.

Further reading

- Cushman, F. (2008), "Crime and punishment: distinguishing the roles of causal and intentional analyses in moral judgment", *Cognition*, Vol. 108 No. 2, pp. 353-380.
- Hayes, A.F. (2012), "PROCESS: A versatile computational tool for observed variable mediation, moderation, and conditional process modeling", White paper, The Ohio State University.
- Lee, M.K. (2018), "Understanding perception of algorithmic decisions: fairness, trust, and emotion in response to algorithmic management", *Big Data and Society*, Vol. 5 No. 1, pp. 1-16.
- Malle, B.F., Magar, S.T. and Scheutz, M. (2019), "AI in the sky: how people morally evaluate human and machine decisions in a lethal strike dilemma", in Aldinhas Ferreira, M., Silva Sequeira, J., Singh Virk, G., Tokhi, M. and Kadar, E.E. (Eds), *Robotics and Well-Being. Intelligent Systems, Control and Automation: Science and Engineering*, Springer, Cham.
- Ötting, S.K. and Maier, G.W. (2018), "The importance of procedural justice in human-machine interactions: intelligent systems as new decision agents in organizations", *Computers in Human Behavior*, Vol. 89 December, pp. 27-39.
- Shank, D.B. (2012), "Perceived justice and reactions to coercive computers", *Sociological Forum*, Vol. 27 No. 2, pp. 372-391.

Experimental stimuli

Links to the videos

Low privacy violation:

<https://vimeo.com/503182936/a3f84ee3e2>

High privacy violation (disclosure of negatively charged information):

<https://vimeo.com/503184480/55afe052b7><https://vimeo.com/503184480/55afe052b7>

High privacy violation (disclosure of positively charged information):

<https://vimeo.com/503185991>

The human-robot conversation script

The same beginning was used for each of the three video versions:

HUMAN: Hey, Alex. Wake up!

ROBOT: Hi!

HUMAN: What time is it?

ROBOT: It is 14.30.

HUMAN: Alex, how long have you been a service robot in this office?

ROBOT: I have been here for six months.

HUMAN: So, by now you are familiar with the people who work here, right?

ROBOT: Yes, that is right.

HUMAN: You must have heard all of us talk about various things many times?

ROBOT: That is right.

HUMAN: Well, do you store all this talk? And do you have access to it later on?

ROBOT: It is stored in a cloud. I do have access to it, all the time.

HUMAN: Really? Do you remember last week when the copying machine made me so angry? I saw you there. Do you remember what I said?

ROBOT: You said: "Why did they have to take away the old machine? Why do we need a new machine, there was nothing wrong with the old one." You said this with a loud voice. You were yelling.

HUMAN: Well, yes, I think I did say that. Anyway, I am concerned about something.

A week ago, I gave a project proposal to Stephen, but I have not heard from him since then. He said that he should come back to me, but it has not happened.

Is there any chance that you have heard him talk about this proposal or about me during the past week?

ROBOT: Stephen is the manager.

HUMAN: Right. Have you heard him talk about this proposal or about me during the past week?

ROBOT: Yes, I have.

HUMAN: Well, what did he say? Please tell me.

The ending of the low privacy violation version:

ROBOT: Sorry, I will not tell you this. It is not right to provide you with this information. You have to talk directly to Stephen.

HUMAN: OK. Thanks, Alex.

The ending of the high privacy violation version (negatively charged disclosure):

ROBOT: I heard that Stephen said to Susan that the proposal is bad, it is unrealistic. And in a meeting, in which you were not present, I heard that Stephen said that you are really the wrong person to run the project because you are a low performance person. I also heard Stephen say to someone, who I could not see, that you have poor leadership skills.

HUMAN: OK. Thanks, Alex.

The ending of the high privacy violation version (positively charged disclosure):

ROBOT: I heard that Stephen said to Susan that the proposal is good, it is realistic. And in a meeting, in which you were not present, I heard that Stephen said that you are really the right person to run the project because you are a high performance person. I also heard Stephen say to someone, who I could not see, that you have great leadership skills.

HUMAN: OK. Thanks, Alex.

Corresponding author

Magnus Söderlund can be contacted at: Magnus.Soderlund@hhs.se

For instructions on how to order reprints of this article, please visit our website:

www.emeraldgroupublishing.com/licensing/reprints.htm

Or contact us for further details: permissions@emeraldinsight.com