Trusting robocop: gender-based effects on trust of an autonomous robot

Sociology, Identity

Robots are becoming omnipresent and use across an increasingly broad portion of society is growing (Breazeal, 2002). Robots are used in entertainment, for stocking shelves and organizing warehouses, for delivering medical equipment, and in other service-related industries (e. g., hospitality). Robots are also being used in security-based contexts to monitor property and protect humans. The Knightscope (K5) robot is one such example as a 152 cm tall, 181. 5 kg mobile robot that autonomously monitors a prescribed area and feeds data back to human data analysts for decision making (Wiggers, 2017). Robots like the Knightscope use many sensors to relay information about suspicious activities and people to their clients. These sensors may have complex machine learning algorithms that analyze massive amounts of data in speeds faster than humans but are essentially opaque to humans, lacking understandability and reducing trust (Christensen & Lyons, 2017). That complexity coupled with the physical size of the system creates an inherent vulnerability to the robot and hence the need to understand the public’s trust in autonomous security robots. Trust of autonomous robots has long been a concern for robots within the military domain as individuals fear the concept of a “ killer robot” (Lyons & Grigsby, 2016). Yet, current robots such as the Knightscope robots are not designed to apprehend or otherwise engage potential criminals. However, this physical engagement may be an option for robots in the future. Thus, it is critical to understand societal perceptions of autonomous security robots that possess the capacity to physically engage (i. e., harm) a human. These perceptions may vary according to individual differences such as gender as males and

females may experience different levels of vulnerability in relation to other beings such as autonomous security robots.

Contemporary research has examined the construct of trust in a human-robot interaction (HRI) context. Trust represents one's willingness to be vulnerable to actions of others (Mayer, Davis, & Schoorman, 1995), and this intention to be vulnerable is pertinent for interpersonal interactions (Colquitt, Scott, & LePine, 2007) and interactions with machines (Lyons, & Stokes, 2012). Key attributes driving one's willingness to be vulnerable stem from two areas: individual differences and trustworthiness perceptions (Mayer et al., 1995). Trustworthiness perceptions are characterized by perceptions of a target's ability (A; i. e., is this person competent?), benevolence (B; does this person have my best interests in mind?), and integrity (I; does this person share my values and are those values relatively stable?). Research has consistently demonstrated that higher levels of A, B, and I are associated with greater willingness to be vulnerable (Colquitt et al., 2007; Mayer & Davis, 1999). Yet, the ABI model originated in the management literature, and thus it is largely unclear how these facets operate in the context of HRI (for a notable exception see Calhoun, Bobko, Gallimore, & Lyons, under review).

A great deal is known about predictors of trust in a human-machine context. Several meta-analyses have been conducted on this topic (see Hancock, Billings, Schaefer, Chen de Visser, & Parasuraman, 2011; Schaefer, Chen, Szalma, & Hancock, 2016) and comprehensive reviews have been written detailing the human-machine trust process (see Hoff & Bashir, 2015, as an

example). One consistent finding in this literature is the fact that individual differences may shape how humans trust machines. This was noted as one of three primary factors (denoted as a “ Human Factor”) influencing the trust process according to the two extant meta-analyses (Hancock et al. 2011; Schaefer et al., 2016) and a recent review article on the topic (Hoff et al., 2015). Thus, trust of machines may be shaped by individual differences.

According to Mayer and colleagues (1995), the primary individual difference factor that influences trust is one’s trait-based trust (i. e., propensity to trust). As a trait, individuals may vary in their general willingness to be vulnerable to others, absent a specific target. Research has shown that one’s propensity to trust has the strongest impact on the trust process when there is little other information available on which to base trust decisions (Alarcon, Lyons, & Christensen, 2016). From this perspective, in the absence of information related to trustworthiness or other socially available information regarding a trust referent, individuals’ reliance decisions may be based on the individual differences that shape how they view and interpret novel stimuli. Recent research has shown that propensity to trust is associated with all aspects of the trust process (beliefs, intentions, and behavior) above and beyond traditional personality measures (see Alarcon, Lyons, Christensen, Capiola, Klosterman, & Bowers, 2018).

Individual differences and their influence on trust in automation have been examined in some prior human factors research. Merritt and Ilgen (2008) found that dispositional trust and extraversion were related to trust in automation, particularly earlier in the interactive process, before the trustor

had established some basis for trustworthiness beliefs. Merritt, Unnerstall, Lee, and Huber (2015) also examined the construct of the perfect automation schema (PAS), one’s trait-based belief regarding the performance of automated systems, and found that components of the PAS were associated with higher trust in automation. Lyons and Guznov (in press) further examined the influence of PAS on trust in automation and found that high expectations (one facet of PAS) were associated with higher trust across three studies. The current study examined one of the fundamental individual differences, namely gender effects on trust of an autonomous security robot.

While not a robotics context, several economic behavioral studies have found gender differences in trust using the Investment Game (Berg, Dickhaut, & McCabe, 1995). This experiment consists of one subject being placed in room A and a second subject placed in room B. Subject A is given $10 and is asked to decide how much of the $10 to send to subject B, knowing that each dollar sent would be tripled once it reached subject B and that subject B would then decide how much money to send back to subject A. The dependent variables are the amount of money sent by subject A and the proportion of money returned by subject B. Some studies found no significant gender differences in trust (Croson & Buchan, 1999; Schwieren & Sutter, 2008), while other studies have shown a tendency for men to exhibit higher levels of trust compared to women (Buchan, Croson, & Solnick, 2008; Chaudhuri & Gangadharan, 2003). However, women were found to be significantly more trustworthy than men, measured by the amount sent back

to subject A (Buchan et al., 2008; Chaudhuri et al., 2003; Croson et al., 1999).

Haselhuhn, Kennedy, Kray, Van Zant, and Schweitzer (2015) modified the Investment Game slightly by informing participants that they would receive $6, which they could then either keep or pass to a counterpart, in which case the money would be tripled. The counterpart could then either keep all of the money or pass half of the money back. What subjects did not know was that their counterpart was computer-simulated. Unlike Berg et al., this study consisted of seven exchange rounds. In rounds one through four, the computer counterpart always returned half of the endowment, whereas in rounds five through six the computer kept all of the money, demonstrating untrustworthy behavior. The seventh round was announced as the final round and was used to measure trust. Results showed that women are more likely to maintain trust after repeated trust violations compared to men (Haselhuhn et al., 2015). The researchers performed a second study where trust violations were committed by the computer-counterpart during rounds one through three in order to examine gender differences in trust recovery. Women were found to show a significantly greater willingness to restore trust after repeated trust violations (Haselhuhn et al., 2015). Thus, females appear to evidence higher trust overall in the context of the Investment Game. However, it is unclear if there are differences when considering robots.

Relatively few studies have examined gender differences in relation to trust in automation and of those that have, findings are inconsistent (Hoff et al.,

2015). However, research examining human interactions with different types of technology has indicated that the communication style and physical appearance of automated systems can moderate (or produce) response-based gender differences (Lee, 2008; Nomura, Kanda, & Suzuki, 2006). Lee (2008) presented both male and female computer aids to male and female participants. Female participants were more influenced by computer-based flattery relative to males, and male-gendered computers elicited greater compliance (Lee, 2008). Research by Nomura and colleagues (2006) reported that females evidence less anxiety associated with robots, relative to males. However, other research has demonstrated that females reported higher anxiety of robots, relative to males (de Graaf & Allouch, 2013). In fact, de Graaf and Allouch (2013) found that female anxiety increased following an interaction with a robot whereas males did not evidence an increase in anxiety following interactions with the robot. Thus, additional research is warranted to help elucidate the role of gender differences in shaping attitudes towards technology. Knowing gender differences in human-robotic trust may aid researchers in the development of autonomous robots that cater to individual differences in trust.

Recently, research in the HRI domain has focused more on physical attributes of the robot, such as its appearance and personality (Powers, Kiesler, & Goetz, 2003; Siegel, Breazeal, & Norton, 2009; Tay, Jung, & Park, 2014; Woods, Dautenhahn, Kaouri, te Boekhorst, & Koay, 2005), as opposed to user attributes such as gender. Unsurprisingly, the perceived gender of a robot plays a role in how men and women interact with, and trust the robot.

Tay et al., (2014) found that participants were more accepting of robots with gender and personalities that conformed to their occupation's general role stereotypes (e. g., male security robots or female healthcare robots). However, perceived trust of the social robots was not influenced by gender-occupational role conformity (Tay et al., 2014). Another study examining the effects of robot gender on human behavior found that participants were more likely to rate the robot of the opposite sex as more credible, trustworthy, and engaging (Siegel et al., 2009). Thus, user attributes are also highly important in the context of HRI. Mutlu, Osman, Forlizzi, Hodgins, and Kiesler (2006) showed a gender effect and its interaction with task structure (cooperative vs. competitive) during an interactive two-player video game played with Honda's Asimo robot. Men found Asimo less desirable in the competitive task compared to the cooperative task whereas women's ratings of desirability did not change across task structure (Mutlu et al., 2006). These findings suggest that men evaluate robots based on the structure of the task being performed, while women make evaluations based off of interactive or social behavior (Mutlu et al., 2006). Studies such as the one performed with Asimo are important for understanding potential user gender differences in human-robot interactions. Examining gender differences will guide the design and implementation of autonomous security robots in different contexts (e. g., hospitals, university campuses) thereby maximizing benefits to users, and minimizing potential risks such as unnecessary harm.

The current study examined gender differences in attitudes toward an autonomous robot that (ostensibly) possessed the capacity to intentionally

harm a human. Based on the above literature, it was expected that females would report 1) higher trust and 2) higher trustworthiness of an autonomous robot relative to males. Attitudes toward the use of the robot in various contexts were also examined. There were no explicit hypotheses with regard to gender for these attitudes and they are reported herein as exploratory analyses to help motivate further investigation in the literature. Reliance intentions

An independent samples t-test was used to compare levels of trust in the robot for males and females. There was a statistically significant difference in levels of trust between males (M = 3. 39, SD = 1. 48) and females (M = 3. 89, SD = 1. 40); t(198) = -2. 36, p = . 019. These results suggest that females were more trusting of the robot.

Trustworthiness

Trustworthiness was measured by assessing participants' perceptions of the robot's ability, benevolence, and integrity. An independent samples t-test was used for each category to compare males and females in their assessment of the robot's trustworthiness. Female (M = 4. 73, SD = 1. 35) perceptions of the robot's ability was statistically greater than male (M = 4. 30, SD = 1. 48) perceptions, t(198) = -2. 04, p < . 05. This was also true regarding female (M = 3. 11, SD = 1. 53) perceptions of the robot's benevolence compared to males (M = 2. 62, SD = 1. 27), t(198) = -2. 43, p < . 05. However, there were no statistically significant differences between

females (M = 4. 18, SD = 1. 19) and males (M = 3. 99, SD = 1. 08) regarding the robot's integrity, t(198) = 1. 19, ns.

Desire to use

As shown in Table 1, females (M = 2. 64, SD = 1. 11) were more likely to want an autonomous security robot in a hospital setting than males (M = 2. 26, SD = 1. 16), t(198) = -2. 24, p < . 05. Results also show that females (M = 2. 78, SD = 1. 21) were more likely than males (M = 2. 44, SD = 1. 20) to want an autonomous security robot on a college campus, t(198) = -1. 96, p = . 051. Lastly, there was a trend showing females (M = 2. 16, SD = 1. 14) were more likely to want a security robot in the home compared to males (M = 1. 90, SD = 0. 99), t(198) = -1. 73, p < . 1. Females and males did not differ on their desire to use an autonomous security robot in the following settings: a military installation, at a forward operating base, in a low-crime neighborhood, in a high-crime neighborhood, at a government building, in a police station, for crowd control at a public social event, or for crowd control at a public military event.

Discussion

As social, interactive robots are being designed for ubiquitous use in homes and organizations (Breazeal, 2002), we must understand societal attitudes towards robots. Much of the contemporary HRI research has focused on the gender of the robot versus the gender of the human (Carpenter, Davis, Erwin-Stewart, Lee, Bransford, & Vye, 2009; Siegel et al., 2009; Tay et al., 2014). Furthermore, outside of the military context, little is known regarding

societal views toward robots that can intentionally engage humans in ways that can be physically harmful. The current study sought to address this research gap by examining gender-based attitudes towards an autonomous robot.

Females were found to be more trusting of the robot compared to males, supporting our first hypothesis. This is consistent with the literature on female anxiety toward robots (Nomura et al., 2006). There are relatively few studies examining user gender differences in HRI contexts, and to these authors' knowledge there are no studies that specifically examined gender differences in trust of autonomous robots that are capable of harming a human. Females report greater perceived risk for crimes and view themselves as less able to physically defend themselves from crime (Jackson, 2009), thus females may have viewed the robot as an objective guard and less likely to exploit females.

While females' (versus males') perceptions of the robot's ability and its benevolence towards others was greater, no gender differences were found for perceptions of integrity. Thus, our second hypothesis was only partially supported. One possibility for this outcome could be related to the security robot being perceived as male (male voice); therefore female participants were more likely to evaluate the robot as more trustworthy, supporting the findings of Siegel and colleagues (2009). Another possibility is that male participants evaluated the robot based off of the task structure, whereas female participants based their judgements on social behavior (Mutlu et al., 2006). When examining the robot through the lens of task

structure/performance, one could argue that there was a failure on the robot’s behalf. However, if social behavior was the main consideration, one could perceive that the security robot behaved in a manner which was socially appropriate, as the robot not only informed the individual to proceed to the main security facility after being denied access, but also gave several warnings when the individual did not listen before using force. This could explain why females found the robot more trustworthy than males, which, notably, contradicts findings of Schermerhorn, Scheutz, and Crowell (2008) who showed females view the robot as more machine-like than human-like compared to males. Therefore, more clarification is warranted.

Surprisingly, there were no differences between males and females with regard to integrity perceptions. It is quite possible that the transient interaction did not allow males or females to establish stable perceptions of integrity. Further, there may have been fewer observable indicators from which participants could have based their integrity perceptions. The robot’s ability and benevolence both had observable indicators for participants to gauge their trustworthiness evaluations, but integrity perceptions may (a) take longer to develop and (b) may require a broader set of observables relative to ability and benevolence. Future research should examine this speculation.

Like trust, females appeared to be more accepting of the robot in certain contexts relative to males. Females were more accepting of the robot on college campuses, in hospitals, and to some extent in one’s home (though this latter finding should be interpreted with caution given its marginal

significance). Females are more likely to perceive risk in public settings as they report greater fear of crime relative to males (Jackson, 2009; Sutton & Farrall, 2005). Females may perceive greater risk overall, and hence may be more accepting of an “ impartial” or objective robotic security provider in public settings such as hospitals and college campuses and within the home.