International Journal of Human - Computer Studies 100 (2017) 48–54

Contents lists available at ScienceDirect

Int. J. Human–Computer Studies

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

Can we control it? Autonomous robots threaten human identity, uniqueness, safety, and resources

Jakub Złotowski[a,*,1], Kumar Yogeeswaran[b,1], Christoph Bartneck[a]

[a] HIT Lab NZ, University of Canterbury, Christchurch, New Zealand
[b] Department of Psychology, University of Canterbury, Christchurch, New Zealand

ARTICLE INFO

Keywords:
Human-robot interaction
Threat
Attitudes
Social acceptance
Autonomous system

ABSTRACT

Emergence of autonomous machines is a hotly debated topic in mass-media. However, previous research has not empirically investigated whether the perceived autonomy of robots affects their social acceptance. In this study we examined the impact of perceived robot autonomy on realistic threats (threats to human jobs, resources and safety) and identity threats (threats to human identity and distinctiveness), attitude toward robots, and support for robotics research. US based participants watched a video of robots performing various tasks – these robots were presented as either autonomous and capable of disregarding human commands or non-autonomous and only capable of following human commands. Participants who watched videos of supposedly autonomous robots perceived robots in general to be significantly more threatening to humans (both realistic and identity threats) than those who watched a video of non-autonomous robots. Furthermore, exposure to autonomous robots evoked stronger negative attitude towards robots in general and more opposition to robotics research than exposure to non-autonomous robots. Both realistic and identity threats mediated the increase in negative attitudes toward robots and opposition to robotics research, although realistic threats were often the stronger mediator of the two. Our findings have practical implications for research on AI and open new questions on the relationship between robot autonomy and their social impact.

1. Introduction

The development of full artificial intelligence could spell the end of the human race – Stephen Hawking (Waugh, 2015).

If I had to guess what the biggest threat to our existence is, it's probably artificial intelligence – Elon Musk (Waugh, 2015).

The above two quotes by famous science and business people about artificial intelligence are examples of the current attitude in popular culture towards autonomous machines. In Western cultures, a future in which humans have to fight against robots that decided to rebel has been a commonly depicted topic in books and movies. In spite of that, the progress of technology enables researchers to create advanced machines that can perform increasing number of tasks autonomously without human control and supervision. These robots are not only limited to factory settings, but have become part of everyday human environments. Self-driving cars and UAVs are examples of such technology.

On the other hand, autonomous machines pose legal and ethical concerns (Calverley, 2006). The ethical aspects regarding the use of autonomous robots are still not well defined (Arkin and Moshkina, 2007). The United Nations annually discusses the use of Lethal Autonomous Weapons Systems during the Convention on Certain Conventional Weapons. Some scientists argue that since autonomous robots in the warfare context cannot be held responsible for their actions, their use is unethical (Sparrow, 2007).

Considering that the development of autonomous machines is a hot topic in mass-media and politics, it is important to conduct empirical research that could help scientists and engineers understand factors that facilitate or hinder the acceptance of robots and other such technology in society. Autonomy is regarded as an important benchmark in HRI (Kahn et al., 2006). It is one of the requirements for a robot to be seen as a moral agent (Sullins, 2011). Moreover, it affects the extent to which people are willing to use a robot (Stafford et al., 2013) or work with it (Weiss et al., 2008). Attributing anthropomorphic characteristics has been also shown to increase trust in autonomous vehicles (Waytz et al., 2014). In addition, autonomy affects blame and credit attributed to a robot and its human interaction

* Corresponding author.
E-mail address: jaz18@uclive.ac.nz (J. Złotowski).
[1] Indicates that both authors contributed equally.

http://dx.doi.org/10.1016/j.ijhcs.2016.12.008
Received 10 December 2015; Received in revised form 25 October 2016; Accepted 27 December 2016
Available online 28 December 2016

partners (Kim and Hinds, 2006).

Previous research in HRI and HCI has investigated the relationship between social acceptance, and autonomous and intelligent technology. Factors, such as a system's transparency (Kim and Hinds, 2006), controllability (Jameson and Schwarzkopf, 2002), trust (Lewandowsky et al., 2000), prior interaction experience (Kirchbuchner et al., 2015), physical contact (Evers et al., 2010) or sharing driving goals (Verberne et al., 2012) play a role in how people perceive autonomous technology. However, all of these studies focused on specific platforms that were able to independently and automatically perform specific tasks. Therefore, the level of autonomy was limited to doing various actions without human input. In comparison with this work, in this paper we focus on the acceptance of robots that can autonomously decide whether to follow or disregard human instructions. The improved capabilities of robots will sooner or later lead to situations in which an autonomous robot must be able to decide whether to follow or disregard human commands to achieve the goal for which it was built. Disregarding a human command does not imply a scenario from science fiction movies where robots turn against humans. Instead, an autonomous robot requested to harm another human or given two conflicting commands by humans will need to determine what to follow. However, our knowledge regarding social acceptance of such technology is currently limited.

Gray and Wegner (2012) showed that the ability of robots to experience, but not become an agent, leads to uncanny feelings towards a robot. Nevertheless, in this paper we will argue that autonomous robots can be also perceived as threatening and induce negative attitudes toward robots in general when people perceive themselves as losing power and control over such technology. Specifically, when robots are perceived to be autonomous, it undermines human perception of dominance or control over robots. By undermining our control or power over robots, they may be perceived as more threatening to human safety, well-being, resources, (i.e. realistic threat) and also to human uniqueness and distinctiveness (i.e. identity threat) which in turn may impact attitudes toward robots and support for robotics research.

1.1. Social power

According to Galinsky et al. (2015) personal social power is equivalent to the concept of autonomy. In psychology, social power is defined as “an asymmetric control over valued resources in a social relationship” (Magee and Galinsky, 2008). This definition could also be applied in the context of HRI as robots are treated as social agents that can engage in social interactions with their users. Indeed, it has been shown that an advice-giving robot that provided highly threatening advice messages to human autonomy invoked feelings of anger and negative thoughts compared to low threat-to-autonomy advice messages (Roubroeks et al., 2010).

However, it is possible that not only in situations of direct threat to human autonomy, robots will encounter reactance from their human interaction partners. A robot that is an autonomous agent may reduce perceived human control over them. This has been proposed by Norman (1994) who suggested that high autonomy of agents can induce negative emotions in users as it makes them feel a lack of control. Since power operates through the need for control (Guinote, 2007), autonomous machines could produce the outcomes known from work on social power in human relationships.

Power has been shown to affect human cognition, motivation, self- and social perception, physiological states, and behavior (Galinsky et al., 2015). It also affects the relative importance of self versus other (Rucker et al., 2011). High power individuals are perceived as more important as a result of possessing more resources and control, while low power people are seen as dependent.

The impact of social power on individual's functioning is a result of social hierarchy being present among cultures and species since it facilitates organization of multiple individuals (Halevy et al., 2011; Van Vugt et al., 2008). Low power states are aversive and people are motivated to gain power (Horwitz, 1958; Worchel et al., 1978). Moreover, people with low-power experience more negative affect than their high-power peers (Berdahl and Martorana, 2006).

Given the diverse effects of social power on human cognition, attitudes, and behavior, it should not be surprising that undermining human's power or control over technology may also be aversive. People want to maintain control in their lives. The events and cognitions that could reduce personal control tend to evoke efforts to restore the perceived control (Landau et al., 2015). In addition, social experiences and environmental conditions that can potentially diminish perceived control of an individual, result in negative arousal (Kobasa, 1979; Tangney et al., 2004; Taylor, 1983; Thompson, 1991). The behavioral inhibition system triggers hypervigilance and anxious arousal when a person is exposed to a threat (Gray and McNaughton, 2000).

Robots are supposed to serve humans and follow their commands. Therefore, people expect to have control and power over them. However, autonomous machines that make decisions on their own threaten that hierarchy. When human wellbeing and existence is affected by unpredictable factors in the environment, such as would be the case of autonomous robots, that would reduce certainty regarding their power to control the environment (Landau et al., 2015). As a result, autonomous robots could be perceived by people as threatening for their control and power over technology. Since power cues often operate nonconsciously (Smith and Galinsky, 2010) and subjective perception of individual's power has stronger impact than objective power, it would be sufficient for people to believe that autonomous robots undermine their control and power over technology which may induce the negative arousal towards these robots. In the present work, we are looking at whether exposure to robots that are seemingly autonomous with no information about whether these are dangerous or beneficial for humans, influences judgments. In particular we investigate the perceived threat and negative attitude toward autonomous robots.

1.2. Realistic and identity threats

Any individual simultaneously belongs to multiple social groups (e.g., ethnic group, national group, religious group, gender, etc.). At the broadest level, people can self-categorize as human beings, where non-human entities (e.g., other animals, plants, or robots) are perceived as outgroups (i.e., group to which we do not belong). In the context of HRI, humans may see robots as an outgroup – despite the fact that robots are non-living. At least in the case of mass-media, humans certainly perceive robots as posing a potential threat to our safety, uniqueness, and survival (e.g., Terminator, Ex-Machina, Blade Runner, i-Robot). Research emerging from the psychology of intergroup relations distinguishes between two distinct sources of threat that can be posed by an out-group: realistic and identity threats (Riek et al., 2006; Stephan et al., 1999). Realistic threat refers to threats to the material resources, safety, and physical wellbeing of the ingroup. In the context of HRI, autonomous robots may be viewed as threatening human jobs, human safety, or wellbeing, thereby posing a realistic threat to humans.

However, a second source of threat involves a more symbolic threat to human identity or distinctiveness. Specifically, identity threats refer to threats to the ingroups' uniqueness, values, and distinctiveness (Riek et al., 2006; Stephan et al., 1999). Much psychological research demonstrates that people are motivated to perceive their own group as positively distinct from others (Tajfel and Turner, 1986). Autonomous robots may not only pose realistic threats to human safety, wellbeing and scarce materials, but also identity threats by blurring the lines between what is perceived to be human and machine (i.e., ‘us’ and ‘them’), as well as what it means to be an autonomous agent that can control its environment.

1.3. Mediating role of threat on prejudice and discrimination

An extant body of work from the psychology of intergroup relations has shown that both realistic and identity threats fuel intergroup prejudice (i.e., negative attitudes), discrimination, and conflict. For example, realistic threats have been shown to be an underlying factor of intergroup prejudice, discrimination, and conflict (Riek et al., 2006; Stephan et al., 1999; LeVine and Campbell, 1972). Similarly, when a group's distinctiveness, uniqueness or identity is threatened, this may also lead to prejudice, discrimination, and intergroup conflict (Jetten et al., 1997, 1998; Yogeeswaran et al., 2012). In the present research, we examine whether exposure to autonomous robots promotes more negative attitudes toward robots because of increased realistic threats and/or identity relevant threats to humans. Similarly, we also examine whether exposure to such allegedly autonomous robots leads people to express greater opposition to funding of robotics research (as a policy outcome) because people feel that robots are a threat to human safety, resources, well-being, as well as human uniqueness and distinctiveness.

1.4. Hypotheses

Based on the above discussion on the relation between autonomy and power we formulated the following hypotheses that will be addressed in this paper:

- H_{1a}: Exposure to autonomous robots will lead people to perceive robots in general as posing more realistic threats than equivalent exposure to non-autonomous robots.
- H_{1b}: Exposure to autonomous robots will lead people to perceive robots in general as posing more identity relevant threats than equivalent exposure to non-autonomous robots.
- H_2: People will express more negative attitudes toward robots and become more opposed to robotics research after being exposed to autonomous robots than non-autonomous robots.
- H_3: Both realistic and identity threats will mediate the relationship between exposure to autonomous robots and negative attitudes toward robots.
- H_4: Both realistic and identity threats will mediate the relationship between exposure to autonomous robots and opposition toward robotics research.

2. Method

2.1. Participants

Two hundred and thirty-nine participants including 90 men and 149 women were recruited using Amazon's online Mechanical Turk (MTurk) in exchange for $2 USD. Of these participants, 63 failed a critical manipulation check or finished in less than 6 min suggesting they had not watched the entire video manipulation (the video itself is over 4 min long making it impossible to finish the entire study within 6 min). The remaining sample therefore included 176 participants (110 females, 66 males). All participants were US based in the age range of 18–73 years (M = 37.76; SD = 11.94).

We chose MTurk platform to recruit participants as MTurk workers were found in previous studies to be more representative for the USA population than either student based studies or Internet recruited samples commonly used in other experiments (Bartneck et al., 2015; Buhrmester et al., 2011; Ramsey et al., 2016). Furthermore, MTurk respondents have similar levels of education and are employed in similar sectors as respondents of other widely used surveys (Huff and Tingley, 2015). In addition, internal motivation rather than monetary reward has been found to be the main reason why US based subjects work on MTurk (Buhrmester et al., 2011). This evidence gave us confidence that our sample is not less representative and diverse than commonly used samples in the field.

2.2. Manipulation

Participants were randomly assigned to watch one of two videos uploaded as a private video on youtube. The videos depicted robots doing a variety of tasks, including cleaning, driving, and looking after the elderly, as well as robots demonstrating their capabilities such as running, jumping, and performing tai-chi. A wide range of robots from industrial machine-like robots, through humanoids to human-like androids were presented in the video. During these visuals, a male American voice documented the recent advances in robotics and the likely future of robotics. All these robots were described as being capable of performing a range of physical and mental tasks such as tennis, weight-lifting, chess, and puzzle solving. However, in one segment of the video, participants were randomly assigned to hear that this new generation of robots are either: (a) capable of making autonomous decisions such as accepting or rejecting human commands depending on their own assessment of a situation; or (b) not autonomous, and only capable of following human commands. Besides this subtle difference in the narration of the video, all other aspects of the video were identical.

2.3. Measures

2.3.1. Realistic threat

Participants completed five items assessing the degree to which they perceived robots to pose a realistic threat to human jobs, resources, and safety. These items were adapted from previous research using ethnic and national groups (Stephan et al., 1999), and have been used in other research in the context of HRI (Yogeeswaran et al., 2016). Sample items included: “The increased use of robots in our everyday life is causing more job loss for humans”, “In the long run, robots pose a direct threat to human safety and wellbeing.” These items were measured on a 7-point scale from strongly disagree (1) to strongly agree (7).

2.3.2. Identity threat

Participants also completed a 5-item measure of the extent to which they perceived robots to pose a threat to human identity and distinctiveness. These items were adapted from previous research using ethnic and national groups (Yogeeswaran and Dasgupta, 2014), and have been used in other HRI research (Yogeeswaran et al., 2016). Sample items included: “Recent advances in robot technology are challenging the very essence of what it means to be human.”, “Technological advancements in the area of robotics is threatening to human uniqueness.” These items were measured on a 7-point scale from strongly disagree (1) to strongly agree (7).

2.3.3. Negative attitudes toward robots

Participants also completed a measure of the NARS (Negative Attitudes toward Robots Scale); (Nomura et al., 2004), a 14-item scale assessing generalized negative feelings toward robots. The measure comprises three subscales assessing: (a) attitudes toward interactions with robots (6 items; e.g. “I would feel very nervous just standing in front of a robot”); (b) attitudes toward the social influence of robots (5 items; e.g. “I feel that in the future society will be dominated by robots”); and (c) attitudes toward emotional interactions with robots (3 items; e.g. “I would feel relaxed talking with robots”).

2.3.4. Support for robotics research

Participants read a brief paragraph detailing the creation of the National Robotics Initiative started by the U.S. government a few years back. After a short description of the initiative, participants were asked to indicate their support for the program using the item: “How much do you support this initiative? ” Participants were also asked to indicate their support for robotics research using the item: “How much do you support the use of tax payer dollars for robotics research? ” Participants

responded to both items using a 10-point Likert-scale from extremely oppose (1) to extremely favor (10).

2.4. Procedure

Participants were recruited via MTurk. Participants first read an information sheet that outlined the purpose of the study and offered consent to participate. Participants initially answered demographic questions pertaining to their gender, age, and ethnicity. Once these were completed, participants were randomly assigned to watch a video clip that either presented a new generation of robots that were allegedly autonomous (i.e. capable of accepting or rejecting human commands based on their own assessment of a situation) or completely non-autonomous. Participants then completed measures of threat (both realistic and identity threats) and negative attitudes toward robots. Finally, participants completed the policy support measure before being probed for suspicion and debriefed.

3. Results

3.1. Mean differences

3.1.1. Realistic threat

A composite measure of realistic threat was created by averaging all 5-items on the scale after ensuring that it had strong internal consistency ($\alpha = .87$). Using this measure, a one-way ANOVA revealed that participants made to believe that robots were autonomous ($M = 4.60$; $SD = 1.33$) perceived robots in general to pose a greater threat to human jobs, resources, and safety than those made to believe these robots were non-autonomous ($M = 4.13$; $SD = 1.44$), $F(1, 174) = 5.28$, $p = .02$, $\eta_p^2 = .03$ (see Table 1).

3.1.2. Identity threat

A composite measure of identity threat was created by averaging all 5-items on the scale after ensuring that it had strong internal consistency ($\alpha = .91$). A similar one-way ANOVA revealed that robots in general were perceived to be significantly more threatening to human uniqueness and identity when they believed that some robots were autonomous and capable of accepting or rejecting human commands ($M = 4.04$; $SD = 1.58$) compared to when they believed these robots were not autonomous ($M = 3.54$; $SD = 1.69$), $F(1, 173) = 4.10$, $p = .04$, $\eta_p^2 = .02$.

3.1.3. Negative attitudes toward robots

A composite for each of the subscales of the NARS was created. Items for the first subscale on negative attitudes toward interactions with robots were combined after establishing that these items showed high internal consistency (6 items; $\alpha = .82$). Similarly, items for the second subscale on negative attitudes toward the social influence of robots were combined after establishing that these items showed high internal consistency (5 items; $\alpha = .78$). Finally, items from the third subscale assessing negative attitudes toward emotional interactions with robots were also combined after establishing these had high internal consistency (3 items; $\alpha = .79$).

We then proceeded to test whether perceived autonomy of robots impacted people's negative attitudes toward them on each of these subscales. A one-way ANOVA first revealed that participants told that new generation robots were capable of autonomous decisions ($M = 2.61$; $SD = 0.85$) showed similar attitudes toward interactions with robots to those told these robots were non-autonomous ($M = 2.46$; $SD = 0.83$), $F(1, 172) = 1.51$, $p = .22$, $\eta_p^2 = .01$. However, participants told that robots were capable of autonomy ($M = 3.34$; $SD = 0.84$) had significantly more negative attitudes toward the social influence of robots than those told that the robots were completely non-autonomous ($M = 3.07$; $SD = 0.97$), $F(1, 173) = 3.89$, $p = .05$, $\eta_p^2 = .02$. Similarly, participants expressed more negative attitudes toward emotional interactions with robots after being told that robots were capable of autonomous decisions ($M = 3.41$; $SD = 0.97$) than when they were told these robots were completely non-autonomous ($M = 3.10$; $SD = 0.93$), $F(1, 174) = 4.37$, $p = .04$, $\eta_p^2 = .03$.

3.1.4. Support for robotics research

Support for robotics research was calculated by combining the two items described earlier after ensuring they had strong internal consistency ($\alpha = .89$). A one-way ANOVA revealed that exposure to robots that were seemingly capable of autonomous decisions ($M = 5.02$; $SD = 2.52$) led participants to express decreased support for robotics research relative to seeing identical robots that were allegedly non-autonomous ($M = 5.85$; $SD = 2.15$), $F(1, 174) = 5.47$, $p = .02$, $\eta_p^2 = .03$

3.2. Mediation analyses

3.2.1. Negative attitudes toward robots

Using Hayes (2013) PROCESS macro, we computed the indirect effect of realistic and identity threats on the relation between perceived robot autonomy on each dependent variable using bias-corrected bootstrapping with 10,000 resamples. Note that if the confidence interval (CI) does not include zero in these analyses, then the effect is considered statistically significant at $p < .05$.

As robot autonomy had a non-significant effect on negative attitudes toward interactions with robots (i.e., NARS-1), we could not test for mediation with the first NARS subscale. However, using the second NARS subscale as a dependent measure, mediation analyses revealed that both realistic threats (indirect coefficient=0.162, SE=0.075, 95% CI [0.024, 0.321]), and identity threats (indirect coefficient=0.156, SE=0.084, 95% CI [0.006, 0.339]) significantly mediated the effect of perceived robot autonomy on negative attitudes toward the social influence of robots (i.e., NARS-2). Even after statistically controlling for the inter-correlations between realistic and identity threats, analyses revealed that both realistic and identity threats simultaneously play unique mediating roles in the effect of robot autonomy on negative attitudes toward the social influence of robots (realistic threat: indirect coefficient=0.081, SE=0.045, 95% CI [0.014, 0.196]; identity threat: indirect coefficient=0.114, SE=0.066, 95% CI [0.009, 0.272]). These findings suggest that when participants are exposed to a new generation of robots that they perceive to be autonomous, then they perceive robots in general to be a threat to human safety, jobs, resources, as well as human uniqueness and distinctiveness; such threats simultaneously drive people to in turn experience more negative attitudes toward the social influence of robots.

A similar set of analyses using negative attitudes toward the emotional interactions with robots (i.e., NARS-3 subscale) revealed that both realistic threats (indirect coefficient=0.095, SE=0.051, 95% CI [0.019, 0.227]) and identity threats (indirect coefficient=0.081, SE=0.047, 95% CI [0.011, 0.205]) significantly mediated the effects of

Table 1
Descriptive statistics and significance testing for each dependent variable.

Measure	Autonomous Robot	Non-Autonomous Robot		
	Mean (SD)	Mean (SD)	*F*- value	*p*-value
Realistic Threat	4.61 (1.33)	4.13 (1.44)	5.28	0.02*
Identity Threat	4.04 (1.58)	3.54 (1.69)	4.10	0.04*
NARS–1	2.61 (0.85)	2.46 (0.83)	1.51	0.22
NARS–2	3.34 (0.84)	3.07 (0.97)	3.89	0.05*
NARS–3	3.41 (0.97)	3.10 (0.93)	4.37	0.04*
Policy Support	5.02 (2.52)	5.85 (2.15)	5.47	0.02*

* indicates mean differences are significant at $p < .05$ level.

robot autonomy on the NARS-3 subscale. However, when we statistically control for the inter-correlations between realistic and identity threats to examine the unique contribution of each as mediators, analyses revealed that only realistic threat significantly mediated the relation between perceived robot autonomy on negative attitudes toward the emotional interactions with robots (realistic threat: indirect coefficient=0.067, SE=0.049, 95% CI [0.003, 0.206]; identity threat: indirect coefficient=0.044, SE=0.039, 95% CI [−0.007, 0.159]). These findings suggest that when participants perceive newer robots to be capable of rejecting human commands (i.e., capable of autonomy), they perceive robots in general to posing both realistic and identity threats to humans that in turn impact attitudes toward emotional interaction with robots. However, realistic threat appear to be the stronger driver in explaining why exposure to autonomous robots leads people to express more negative attitudes toward emotional interactions with robots.

3.2.2. Support for robotics research

Similar to the analyses above, Hayes (2013) PROCESS macro was used to examine the mediating role of both identity and realistic threats on support and opposition for robotics research. Analyses revealed that both realistic threats (indirect coefficient=−0.418, SE=0.194, 95% CI [−0.838, −0.068]) and identity threats (indirect coefficient=−0.279, SE=0.152, 95% CI [−0.624, −0.025]) mediated the effect of perceived robot autonomy on support for robotics research. Controlling for the relation between realistic and identity threats, analyses revealed that only realistic threat significantly mediated the effect of perceived robot autonomy on support for robotics research (realistic threat: indirect coefficient=−0.409, SE=0.195, 95% CI [−0.875, −0.089]; identity threat: indirect coefficient=−0.055, SE=0.090, 95% CI [−0.310, 0.075]). These findings suggest that when participants are exposed to robots that are allegedly autonomous and capable of rejecting human commands, they perceive robots to pose realistic and identity threats to humans which in turn leads them to oppose robotics research. However, threat to human safety, jobs, and resources (i.e. realistic threat) is the more powerful driver in explaining why exposure to autonomous robots lead perceivers" to oppose robotics research.

4. Discussion

In this study we investigated the impact of robots' autonomy on their perceived threat and attitudes towards them. In particular we hypothesized that autonomous robots will be perceived as more threatening and evoke more negative reactions than non-autonomous robots.

The results supported our H_{1a} – after watching a video of presumably autonomous robots, participants perceived robots as posing more realistic threat than if the robots were non-autonomous. Since this measure is related with material threats, autonomous robots could be perceived as posing higher risk for human jobs that they could take over. Moreover, this threat also indicates that people are concerned about their safety. Considering that our participants were US based, Hollywood pop-culture that presents robots with advanced AI as rising against human kind could be responsible for the results. Since participants were told in the autonomous condition that these latest robots can obey or disregard human commands, it is possible that participants feared that these robots could decide to stand against humans if they disagree with them.

Our H_{1b} – autonomous robots will be perceived as posing more identity threat than non-autonomous robots – was also supported by the results. Identity threat is related with group uniqueness, values and distinctiveness. Autonomous robots appear to blur the line between what is human and what is machine. By showing some form of intentionality and not blindly obeying human commands, they are seen as expressing a core aspect of human uniqueness relative to machines, which is especially threatening to people.

Put together, these results indicate that autonomous robots are more threatening to people than non-autonomous robots. This finding does not support previous research on the uncanny valley that suggested that only the ability of robots to experience, but not to be an agent (independently execute actions) makes them unnerving (Gray and Wegner, 2012). Our results suggest that a robot that can control its behavior can be perceived as threatening. It is possible that participants in Gray and Wegner (2012) study did not perceive the machine as fully autonomous, since they were only told that it can independently execute actions or self-control. In our experiment we explicitly stated that the newest generation of the robots can disregard human commands and therefore we emphasized robot autonomy more explicitly. Moreover, machines that can independently execute actions already exist. On the other hand, robots that can disregard human commands are more abstract. Since people have especially strong negative attitudes when faced with unfamiliar scenarios for HRI (Enz et al., 2011) it is possible that our scenario is more unfamiliar than the one presented by Gray and Wegner (2012).

We found partial support for H_2 – autonomous robots will invoke more negative attitude and lead people to become more opposed to robotics research than non-autonomous robots. Our results show that people had more negative attitude toward social influence of robots and emotional interactions with robots after watching a video of autonomous than non-autonomous robots. However, there was no statistically significant difference between these two conditions for attitudes toward interactions with robots. The effect of robot autonomy on attitudes toward the social influence and emotional interactions with robots could be a result of the fear that if robots gained power and autonomy, they would threaten the established hierarchy. Since social hierarchy is widespread among cultures (Halevy et al., 2011; Van Vugt et al., 2008) an autonomous robot could pose a risk that it might seek higher status in the future (i.e. we anthropomorphize and project characteristics of ourselves onto robots). As low power state is aversive (Horwitz, 1958; Worchel et al., 1978), people's attitudes towards the threat may become more negative.

Participants did not express negative attitudes toward interacting with robots after being exposed to autonomous robots. However, in case of emotional interaction, their attitudes were more negative. It is possible that autonomous robots can be perceived as beneficial for everyday human tasks and when used for work purposes. However, people might be reluctant to trust robots and establish closer social relationships with them following exposure to autonomous robots. People may fear that a robot that can decide on its own whether to follow human instructions, could use human emotions for its own purposes. Such a robot could understand human emotions, but would not necessarily need to be affected by them. Therefore, it could engage in immoral actions for its own benefits. An example of such behavior can be found in the movie 'Ex-machina' where fully autonomous android plays on human emotions and needs to achieve its own goals and obtain freedom at the cost of human life.

People were less willing to support robotics research after being exposed to a video of autonomous robots than non-autonomous robots. Therefore, our findings also have practical implications for the researchers working on AI and autonomous systems. Presenting a robotic platform as potentially capable of becoming autonomous in the future might reduce general public's willingness to support it and in case of crowdfunding campaigns and tax payer driven research, result in less funding for robotic prototypes.

The results of our study partially support our H_3 for 2 NARS subscales that were affected by our manipulation – both realistic and identity threats mediate the relationship between exposure to autonomous robots and negative attitudes toward robots. Participants who were exposed to robots that are capable of rejecting human commands experienced greater threats to human safety and job, as well as human uniqueness, which in turn lead to more negative attitude regarding the social influence of robots in general. A fully autonomous robot not only

blurs the line between what is human and what is machine, but also poses a threat to social hierarchy on which societies are built (Halevy et al., 2011; Van Vugt et al., 2008), which results in negative perceptions of robots in general. Moreover, autonomous robots may be perceived as lacking predictability which would be an additional hazard for human safety. However, in case of negative attitudes toward emotional interaction with robots, when we controlled for the relation between realistic and identity threats, only the former was a significant mediator. Therefore, people had stronger negative attitudes toward emotional interaction with robots after being exposed to allegedly autonomous robots mainly due to their higher perceived threat to human safety and jobs.

Our H_4 – both realistic and identity threats will mediate the relationship between exposure to autonomous robots and opposition toward robotics research – was partially supported by the results. Participants who were exposed to robots that are capable of rejecting human commands invoked threats to human safety and jobs, as well as human uniqueness, which in turn lead to reduced willingness to support robotics research. Since participants found autonomous robots to be threatening, it is not surprising that they are also unwilling to support research involving robots as that could lead to development of machines that they fear. However, when controlling for the relation between realistic and identity threats, only the former is a significant mediator. This suggests that threats to human safety, job, and resources are the major driver behind why exposure to autonomous robots leads people to oppose robotics research.

Overall, the results clearly show that people have ongoing fears of autonomous robots. In our study we did not make any claims regarding robots potentially being good or bad, or introduce any scary stories about a future with autonomous robots. Only 3 sentences in over 4 min long video (less than 10% of video duration time) emphasized the robots' autonomy. If this information was presented more saliently, the effects could have been even stronger. Our manipulation only triggered existing fears as there is no logical reason why an autonomous agent has to be evil.

4.1. Limitations & future work

In this study all participants were US based. Therefore, our results could not be generalized to other cultures and could be culture specific. Autonomous robots are still a thing of the future and cannot be found in everyday human environments. Furthermore, general public lacks knowledge to understand how AI works. Therefore, people have to base their opinions on other sources of information. It is possible that their perception of autonomous robots is strongly and predominantly influenced by the mass-media, books and movies. In Western cultures, the media typically present robots as rising against humans and ending in a conflict. However, that is not the case for other cultures. For example, in East Asia, robots are presented as helping humans to defeat common enemies and they can co-exist peacefully (Kaplan, 2004). Therefore, it is possible that the perception of autonomous robots in these cultures may be much more positive than in our study.

In this study we used a video material of robots rather than having a live HRI. Although, the effects could be stronger when people are placed in front of a robot that is supposedly autonomous, the present research suggests that even a video clip about autonomous robots can produce statistically significant effects. Using a video allowed us also to include multiple robots that varied in their appearance, skills and human-likeness, as it is not practical to run a study with more than few physically present robots.

On the other hand, interaction with autonomous robots could help to alleviate the fears. The factors that helped to affect the acceptance of intelligent technology (Kim and Hinds, 2006; Kirchbuchner et al., 2015; Verberne et al., 2012) may be equally suitable for decreasing perceived threats of autonomous robots. A robot that can explain why it disregards a human command or makes a specific decision, may be perceived as less threatening than a robot who makes the decision for unknown reasons.

In the introduction we linked robots' autonomy with a threat of losing social power by humans. Although our results are consistent with what could be hypothesized from the link between power and autonomy, this study did not explicitly measure loss of social power. Therefore, future research should measure social power in order to establish its role as a potential mediator of the presented findings.

Future research could also incorporate elements of personality and individual difference measures to examine the potential moderating role of such factors on people's reaction to autonomous robots. For example, people high in threat sensitivity or belief in dangerous world may be especially likely to respond negatively to our manipulations. Similarly, it is possible that realistic and identity threats posed by autonomous robots have different importance for different groups of people. Identity threat could be universal and experienced by all people since it concerns humankind uniqueness and distinctiveness as a whole. On the other hand, realistic threat is resource and safety specific. It is possible that people who already have a low status in social hierarchy will experience higher realistic threat by autonomous robots than high status individuals. Firstly, autonomous robots will take simple and unqualified jobs, that low power individuals possess. High power individuals, at least in the first stages, will benefit from introduction of autonomous machines as they may reduce the working costs in their businesses. On the other hand, high power individuals may be especially sensitive to threats to their current standing which may lead them to react especially strongly to autonomous robots. The present research offers just a starting point for many such future explorations.

Although, MTurk workers are employed in diverse industries (Huff and Tingley, 2015), it is possible that participants in our study represented professions that are especially endangered by the introduction of robots. Therefore, they might have reported higher realistic threat than the general US population. Thus, if general public experiences lower level of realistic threat than our sample, the role of identity threat would have been stronger. Future work could evaluate the impact of income and profession on these threats.

5. Conclusions

Research on AI and autonomous robots is not going to stop in the near future. However, its speed is partially affected by the available funding and acceptance of autonomous robots by the general public. Therefore, it is important to understand how people perceive autonomous machines and what affects that perception. In this study we investigated the impact of perceived robots' autonomy on the threat posed by them and attitude towards robots in general. We found that robots perceived to be autonomous increase realistic and identity threats, and evoke negative attitudes toward them compared with non-autonomous robots. Furthermore, people are more reluctant to support robotics research after watching a video of supposedly autonomous robots. Both realistic and identity threats mediate the effects of autonomous robots on attitude toward robots in general and willingness to support robotic research, with realistic threat being a stronger mediator on attitude toward emotional interactions with robots and support for robotic research.

The presented study was the first step to investigate the relationship between robots' autonomy and their social acceptance. Future research should focus on establishing why autonomous robots are perceived as threatening and evoke negative attitudes, and how to increase their acceptance. In particular, we proposed that potential loss of power over autonomous machines may be responsible for their low social acceptance. Understanding the psychological processes behind our results can facilitate development of AI that will be socially acceptable and can benefit the community by increasing the support of general public for such research.

References

Arkin, R.C., Moshkina, L., 2007. Lethality and autonomous robots: An ethical stance. In: Proceedings of the 2007 International Symposium on Technology and Society: Risk, Uncertainty, Vulnerability, Technology and Society, ISTAS, June 1, 2007 - June 2, 2007 International Symposium on Technology and Society, Institute of Electrical and Electronics Engineers Inc., ⟨http://dx.doi.org/10.1109/ISTAS.2007.4362202⟩.

Bartneck, C., Duenser, A., Moltchanova, E., Zawieska, K., 2015. Comparing the similarity of responses received from studies in amazons mechanical turk to studies conducted online and with direct recruitment. PLoS One 10, 1–23. http://dx.doi.org/10.1371/journal.pone.0121595.

Berdahl, J.L., Martorana, P., 2006. Effects of power on emotion and expression during a controversial group discussion. Eur. J. Soc. Psychol. 36, 497–509. http://dx.doi.org/10.1002/ejsp.354.

Buhrmester, M., Kwang, T., Gosling, S.D., 2011. Amazon's mechanical turk a new source of inexpensive, yet high-quality, data? Perspect. Psychol. Sci. 6, 3–5. http://dx.doi.org/10.1177/1745691610393980, (arXiv:http://pps.sagepub.com/content/6/1/3.full.pdf+html.).

Calverley, D.J., 2006. Android science and animal rights, does an analogy exist? Connect. Sci. 18, 403–417.

Enz, S., Diruf, M., Spielhagen, C., Zoll, C., Vargas, P.A., 2011. The social role of robots in the future-explorative measurement of hopes and fears. Int. J. Soc. Robot. 3, 263–271. http://dx.doi.org/10.1007/s12369-011-0094-y.

Evers, V., Winterboer, A., Pavlin, G., Groen, F., 2010. The evaluation of empathy, autonomy and touch to inform the design of an environmental monitoring robot. In Ge, S.S., Li, H., Cabibihan, J.-J., Tan, Y.K. (Eds.), Social Robotics In: Proceedings of the Second International Conference on Social Robotics, ICSR 2010, Singapore, November 23-24, 2010, pp. 285–294. Berlin, Heidelberg:Springer Berlin Heidelberg. ⟨http://dx.doi.org/10.1007/978-3-642-17248-9_30⟩

Galinsky, A.D., Rucker, D.D., Magee, J.C., 2015. Power: Past findings, present considerations, and future directions. In Mikulincer, M., Shaver, P.R., Simpson, J.A., Dovidio, J.F. (Eds.), APA handbook of personality and social psychology, vol. 3: Interpersonal relations APA handbooks in psychology. Washington, DC, US:American Psychological Association, pp. 421–460.

Gray, J.A., McNaughton, N., 2000. The Neuropsychology of Anxiety: an Enquiry Into the Function of the Septo-hippocampal System 2nd ed.,. Oxford University Press, New York, NY.

Gray, K., Wegner, D., 2012. Feeling robots and human zombies mind perception and the uncanny valley. Cognition 125, 125–130.

Guinote, A., 2007. Power affects basic cognition: Increased attentional inhibition and flexibility. J. Exp. Soc. Psychol. 43, 685–697. http://dx.doi.org/10.1016/j.jesp.2006.06.008.

Halevy, N., Chou, E.Y., Galinsky, A.D., 2011. A functional model of hierarchy why, how, and when vertical differentiation enhances group performance. Organ. Psychol. Rev. 1, 32–52. http://dx.doi.org/10.1177/2041386610380991.

Hayes, A.F., 2013. Introduction to Mediation, Moderation, and Conditional Process Analysis: a Regression-Based Approach. Guilford Press, New York, NY.

Horwitz, M., 1958. The veridicality of liking and dis- liking. In: Tagiuri, R., Petrullo, L. (Eds.), Person Perception and Interpersonal Behavior. Stanford University Press, Stanford, CA, 165–183.

Huff, C., Tingley, D., 2015. “Who are these people?” evaluating the demographic characteristics and political preferences of mturk survey respondents. Res. Polit., 2. http://dx.doi.org/10.1177/2053168015604648.

Jameson, A., Schwarzkopf, E., 2002. Pros and cons of controllability: An empirical study. In De Bra, P., Brusilovsky, P., Conejo, R., (Eds.), Adaptive Hypermedia and Adaptive Web-Based Systems In: Proceedings of the Second International Conference, AH 2002 Málaga, Spain, May 29–31, 2002, pp. 193–202. Berlin, Heidelberg:Springer Berlin Heidelberg. ⟨http://dx.doi.org/10.1007/3-540-47952-X_21⟩.

Jetten, J., Spears, R., Manstead, A.S., 1997. Distinctiveness threat and prototypicality combined effects on intergroup discrimination and collective self-esteem. Eur. J. Soc. Psychol. 27, 635–657.

Jetten, J., Spears, R., Manstead, A.S., 1998. Defining dimensions of distinctiveness group variability makes a difference to differentiation. J. Personal. Soc. Psychol. 74, 1481–1492. http://dx.doi.org/10.1037/0022-3514.74.6.1481.

Kahn, P., Ishiguro, H., Friedman, B., Kanda, T., 2006. What is a Human? - Toward Psychological Benchmarks in the Field of Human-Robot Interaction. In: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006, pp. 364–371. ⟨http://dx.doi.org/10.1109/ROMAN.2006.314461⟩.

Kaplan, F., 2004. Who is afraid of the humanoid? Investigating cultural differences in the acceptance of robots. Int. J. Hum. Robot. 01, 465–480. http://dx.doi.org/10.1142/S0219843604000289.

Kim, T., Hinds, P., 2006. Who Should I Blame? Effects of Autonomy and Transparency on Attributions in Human-Robot Interaction. In: Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication, 2006. ROMAN 2006 pp. 80–85. ⟨http://dx.doi.org/10.1109/ROMAN.2006.314398⟩.

Kirchbuchner, F., Grosse-Puppendahl, T., Hastall, M.R., Distler, M., Kuijper, A., 2015. Ambient intelligence from senior citizens' perspectives: Understanding privacy concerns, technology acceptance, and expectations. In B. De Ruyter, A. Kameas, P. Chatzimisios, & I. Mavrommati (Eds.), Ambient Intelligence In: Proceedings of the 12th European Conference, AmI 2015, Athens, Greece, November 11-13, 2015, pp. 48–59. Cham: Springer International Publishing. ⟨http://dx.doi.org/10.1007/978-3-319-26005-1_4⟩.

Kobasa, S.C., 1979. Stressful life events, personality, and health an inquiry into hardiness. J. Personal. Soc. Psychol. 37, 1–11. http://dx.doi.org/10.1037/0022-3514.37.1.1.

Landau, M.J., Kay, A.C., Whitson, J.A., 2015. Compensatory control and the appeal of a structured world. Psychol. Bull. 141, 694–722. http://dx.doi.org/10.1037/a0038703.

LeVine, R.A., Campbell, D.T., 1972. Ethnocentrism: Theories of Conflict, Ethnic Attitudes, and Group Behavior volume ix. John Wiley & Sons, Oxford, England.

Lewandowsky, S., Mundy, M., Tan, G., 2000. The dynamics of trust comparing humans to automation. J. Exp. Psychol.: Appl. 6, 104–123.

Magee, J.C., Galinsky, A.D., 2008. Social Hierarchy the self-reinforcing nature of power and status. Acad. Manag. Ann. 2, 351–398. http://dx.doi.org/10.1080/19416520802211628.

Nomura, T., Kanda, T., Suzuki, T., Kato, K., 2004. Psychology in human-robot communication: An attempt through investigation of negative attitudes and anxiety toward robots. In: Proceedings of the 13th IEEE International Workshop on Robot and Human Interactive Communication pp. 35–40.

Norman, D.A., 1994. How might people interact with agents. Commun. ACM 37, 68–71. http://dx.doi.org/10.1145/176789.176796.

Ramsey, S.R., Thompson, K.L., McKenzie, M., Rosenbaum, A., 2016. Psychological research in the internet age the quality of web-based data. Comput. Hum. Behav. 58, 354–360. http://dx.doi.org/10.1016/j.chb.2015.12.049.

Riek, B.M., Mania, E.W., Gaertner, S.L., 2006. Intergroup threat and outgroup attitudes a meta-analytic review. Personal. Soc. Psychol. Rev. 10, 336–353. http://dx.doi.org/10.1207/s15327957pspr1004_4.

Roubroeks, M., Ham, J., Midden, C., 2010. The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals. In: Proceedings of the 5th International Conference on Persuasive Technology, PERSUASIVE 2010, June 7, 2010 - June 10, 2010 pp. 174–184. Springer Verlag, vol. 6137, LNCS of Lecture Notes in Computer Science. ⟨http://dx.doi.org/10.1007/978-3-642-13226-1_18⟩.

Rucker, D.D., Dubois, D., Galinsky, A.D., 2011. Generous paupers and stingy princes power drives consumer spending on self versus others. J. Consum. Res. 37, 1015–1029. http://dx.doi.org/10.1086/657162.

Smith, P.K., Galinsky, A.D., 2010. The nonconscious nature of power cues and consequences. Soc. Personal. Psychol. Compass 4, 918–938. http://dx.doi.org/10.1111/j.1751-9004.2010.00300.x.

Sparrow, R., 2007. Killer robots. J. Appl. Philos. 24, 62–77. http://dx.doi.org/10.1111/j.1468-5930.2007.00346.x.

Stafford, R.Q., MacDonald, B.A., Jayawardena, C., Wegner, D.M., Broadbent, E., 2013. Does the robot have a mind? Mind perception and attitudes towards robots predict use of an eldercare robot. Int. J. Soc. Robot. 6, 17–32. http://dx.doi.org/10.1007/s12369-013-0186-y.

Stephan, W.G., Ybarra, O., Bachman, G., 1999. Prejudice toward immigrants. J. Appl. Soc. Psychol. 29, 2221–2237. http://dx.doi.org/10.1111/j.1559-1816.1999.tb00107.x.

Sullins, J.P., 2011. When is a robot a moral agent? In: Anderson, M., Anderson, S.L. (Eds.), Machine Ethics. Cambridge University Press, 151–162.

Tajfel, H., Turner, J., 1986. The social identity theory of intergroup behavior. Psychology of Intergroup Relations, Chicago, IL, USA: Nelson-Hall, pp. 7–25.

Tangney, J.P., Baumeister, R.F., Boone, A.L., 2004. High self-control predicts good adjustment, less pathology, better grades, and interpersonal success. J. Personal. 72, 271–324. http://dx.doi.org/10.1111/j.0022-3506.2004.00263.x.

Taylor, S.E., 1983. Adjustment to threatening events a theory of cognitive adaptation. Am. Psychol. 38, 1161–1173. http://dx.doi.org/10.1037/0003-066X.38.11.1161.

Thompson, S.C., 1991. Intervening to enhance perceptions of control. In Snyder, C.R., Forsyth, D.R., (Eds.), Handbook of social and clinical psychology: The health perspective Pergamon general psychology series, vol. 162. Elmsford, NY, US:Pergamon Press, pp. 607–623. .

Van Vugt, M., Hogan, R., Kaiser, R.B., 2008. Leadership, followership, and evolution some lessons from the past. Am. Psychol. 63, 182–196. http://dx.doi.org/10.1037/0003-066X.63.3.182.

Verberne, F.M., Ham, J., Midden, C.J., 2012. Trust in smart systems sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. Hum. Factor.: J. Hum. Factor. Ergon. Soc. 54, 799–810.

Waugh, R., 2015. Stephen Hawking warns of the dangers of intelligent robots. Metro. URL:⟨http://metro.co.uk/2015/01/13/stephen-hawking-warns-of-the-dangers-of-intelligent-robots-5020270/⟩.

Waytz, A., Heafner, J., Epley, N., 2014. The mind in the machine anthropomorphism increases trust in an autonomous vehicle. J. Exp. Soc. Psychol. 52, 113–117. http://dx.doi.org/10.1016/j.jesp.2014.01.005.

Weiss, A., Wurhofer, D., Lankes, M., Tscheligi, M., 2008. Autonomous vs. tele-operated: How people perceive human-robot collaboration with HRP-2. In: Proceedings of the 4th ACM/IEEE International Conference on Human - Robot Interaction, HRI'09 Association for Computing Machinery, pp. 257–258.. ⟨http://dx.doi.org/10.1145/1514095.1514164⟩.

Worchel, S., Arnold, S.E., Harrison, W., 1978. Aggression and power restoration the effects of identifiability and timing on aggressive behavior. J. Exp. Soc. Psychol. 14, 43–52. http://dx.doi.org/10.1016/0022-1031(78)90059-8.

Yogeeswaran, K., Dasgupta, N., 2014. The devil is in the details abstract versus concrete construals of multiculturalism differentially impact intergroup relations. J. Personal. Soc. Psychol. 106, 772–789. http://dx.doi.org/10.1037/a0035830.

Yogeeswaran, K., Dasgupta, N., Gomez, C., 2012. A new American dilemma? The effect of ethnic identification and public service on the national inclusion of ethnic minorities. Eur. J. Soc. Psychol. 42, 691–705. http://dx.doi.org/10.1002/ejsp.1894.

Yogeeswaran, K., Złotowski, J., Livingstone, M., Bartneck, C., Sumioka, H., Ishiguro, H., 2016. The interactive effects of robot anthropomorphism and robot ability on perceived threat and support for robotics research. J. Hum.-Robot Interact. 5, 29–47 . http://dx.doi.org/10.5898/JHRI.5.2.Yogeeswaran.