Ergonomics

ISSN: 0014-0139 (Print) 1366-5847 (Online) Journal homepage: https://www.tandfonline.com/loi/terg20

Factors that affect younger and older adults' causal attributions of robot behaviour

Richard Pak, Jessica J. Crumley-Branyon, Ewart J. de Visser & Ericka Rovira

ERGONOMICS
https://doi.org/10.1080/00140139.2020.1734242

Taylor & Francis
Taylor & Francis Group

ARTICLE

Factors that affect younger and older adults' causal attributions of robot behaviour

Richard Pak[a], Jessica J. Crumley-Branyon[a], Ewart J. de Visser[b] and Ericka Rovira[c]

[a]Department of Psychology, Clemson University, Clemson, SC, USA; [b]Department of Behavioral Sciences and Leadership, Warfighter Effectiveness Research Center, U. S. Air Force Academy, Colorado Springs, CO, USA; [c]Department of Behavioral Sciences and Leadership, U.S. Military Academy, West Point, NY, USA

ABSTRACT

Stereotypes are cognitive shortcuts that facilitate efficient social judgments about others. Just as causal attributions affect perceptions of people, they may similarly affect perceptions of technology, particularly anthropomorphic technology such as robots. In a scenario-based study, younger and older adults judged the performance and capability of an anthropomorphised robot that appeared young or old. In some cases, the robot successfully performed a task while at other times it failed. Results showed that older adult participants were more susceptible to aging stereotypes as indicated by trust. In addition, both younger and older adult participants succumbed to aging stereotypes when measuring perceived capability of the robots. Finally, a summary of causal reasoning results showed that our participants may have applied aging stereotypes to older-appearing robots: they were most likely to give credit to a properly functioning robot when it appeared *young* and performed a *cognitive* task. Our results tentatively suggest that human theories of social cognition do not wholly translate to technology-based contexts and that future work may elaborate on these findings.

Practitioner summary: Perception and expectations of the capabilities of robots may influence whether users accept and use them, especially older users. The current results suggest that care must be taken in the design of these robots as users may stereotype them.

Abbreviations: HRI: human-robot interaction; M: mean; SD: standard deviation; CPRS: complacency potential rating scale; ANOVA: analysis of variance

ARTICLE HISTORY
Received 30 October 2019
Accepted 21 January 2020

KEYWORDS
Aging; stereotype; anthropomorphism; causal reasoning; human-robot interaction

1. Introduction

People attribute human-like qualities such as personality, mindfulness, and social characteristics to inanimate objects such as technology (Reeves & Nass, 1996; Nass & Moon, 2000). This makes technology susceptible to stereotyping based on its appearance and etiquette (Nass & Lee, 2001; Parasuraman & Miller, 2004; Eyssel & Kuchenbrandt, 2012; Hayes & Miller 2010). For example, when an anthropomorphised computerised aid was included in a tutoring task about computers and technology (a stereotypically male field), participants were more likely to rate a male-appearing aid as more competent and trustworthy than a female-appearing aid (Nass, Steuer, & Tauber, 1994; Nass, Moon, & Green, 1997). In this example, pre-existing, pervasive gender-based stereotypes dictated judgments about the capabilities and trustworthiness of the computerised aid.

The goal of the current study was to examine the extent to which aging stereotypic thinking was activated in younger and older participants by the physical appearance of robots. Inspired by the approach taken by Nass and colleagues (Nass, Steuer, & Tauber, 1994), this study aimed to test existing social cognition models of aging stereotypes as they relate to causal attributions in the context of robotics. It is appropriate to apply well-established theories and methods of social science in an interdisciplinary context (Czaja & Sharit, 2003). In this case, the goal was to understand the implications of aging perceptions to be able to predict how younger and older adults will behave in real-world practical and novel interactions. Although the social cognition literature predicts specific patterns of aging stereotype activation for both younger and older adults, it is unclear whether the construct will directly apply to a new technology

CONTACT Richard Pak richpak@clemson.edu Department of Psychology, Clemson University, Clemson, SC, USA

domain or whether similar patterns will be observed in a human-robot context.

The theoretical relevance of this work is that the results of this study will inform the limits of stereotypic thinking by investigating whether younger and/or older adults apply stereotypes to robots. The practical relevance of this work is that the current study may inform the design of robots to enhance human-robot interaction (HRI), particularly for older adults who tend to be less accepting of technological aids than other age groups (Czaja et al., 2006).

1.1. Stereotypes and aging

Stereotypes are cognitive shortcuts that facilitate efficient social judgments about others (e.g. Ashmore & Del Boca, 1981). There are certain contexts in which stereotypes are more likely to be activated than others. Stereotypes are more likely to be activated in situations where the actor is in a role that is inconsistent with prescriptive societal gender or age roles (e.g. Kuchenbrandt, Häring, Eichberg, Eyssel, & André, 2014). For example, individuals perceived a female-voiced computer to be more informative about romantic relationships than the male-voiced computer (Nass, Moon, & Green, 1997). Physical appearance has been identified as a critical factor in the activation of aging stereotypes (Brewer & Lui, 1984; Hummert, 1994; Hummert, Garstka, & Shaner, 1997; Palumbo, Adams, Hess, Kleck, & Zebrowitz, 2017). In particular, facial features are considered to be a main source of information used to activate stereotypes (Hummert, Garstka, & Shaner, 1997; Kaufmann, Krings, & Sczesny, 2016), suggesting that individuals utilise physical cues to make social judgments. Age group membership also influences perceptions and processing of facial features (Wiese, Komes, & Schweinberger, 2013). Because stereotype activation is relatively automatic, requires few cognitive resources and requires greater cognitive effort to override them (Sherman, 2001), older adults may be more likely than younger adults to apply stereotypes when they do not have other sources of information available to them under ambiguous situations (Bargh, 1994; Von Hippel, Silver, & Lynch, 2000). Adults of all ages exhibit an in-group bias (i.e. own-age bias) within the domains of facial recognition, feature encoding, and estimation of age (Schaich, Obermeyer, Kolling, & Knopf, 2016; Campbell, Murray, Atkinson, & Ruffman, 2017). This suggests that older adults are more likely to accurately recognise and categorise older adult faces as members of their ingroup, which would likely activate age stereotypes more readily within older adults compared to younger adults who view older adults as the outgroup.

1.2. Causal attributions and aging

One prominent way in which stereotypes affect humans is in the types of causal attributions that they make about the performance of others (Fiske & Taylor, 1991). When trying to determine the causality of an event (e.g. blame), people tend to use two types of information: *dispositional* qualities of the individuals involved in an outcome and the influences of the *situation* itself (Gilbert & Malone, 1995; Krull, 1993; Krull & Erickson, 1995). Potential biases in the causal attribution process can come from the valence of the situational outcome (was the outcome positive or negative), the degree of informational ambiguity of the situation, and the degree of control an actor has over an outcome (Blanchard-Fields, 1994; Tomlinson & Mayer, 2009). Blanchard-Fields suggested that, in general, older adults are most likely to make dispositional rather than situational attributions; that is, to attribute the outcome to qualities of the person, when the valence of the outcome of a situation was negative and the actor's role in the outcome was ambiguous. When personal beliefs about another individual or situation are violated (i.e. against prevailing social norms), older adults are also more likely to make dispositional attributions of blame rather than situational (Blanchard-Fields, 1996; Blanchard-Fields, Hertzog, & Horhota, 2012). Thus, we predicted that there would be a main effect of participant age, robot age, and reliability on causal attributions. Our measurement of causal attributions (described further in Section 2.2.3) involves separately measuring the extent to which one makes a situational or dispositional attribution upon viewing an event. Consistent with the social cognition literature, we expect to find that older adults will make significantly higher dispositional attributions than younger adult participants, participants would make higher dispositional attributions when the robot appears older, and dispositional ratings would be higher for unreliable task performance than reliable task performance. This is because older adults are more likely to make dispositional (i.e. internal) attributions of blame when an outcome of an event is perceived as negative (the unreliable condition) and when their beliefs are violated (i.e. when an older looking robot performs the cognitive and physical tasks; Blanchard-Fields, Hertzog, & Horhota, 2012).

1.3. Perceived capabilities and aging

Stereotypes also influence our expectations or predictions of the capabilities of the subject of the stereotype. Because adults of all ages expect memory performance to decline with age (Lineweaver & Hertzog, 1998), older adults' memory capabilities are *perceived* more negatively than other age groups (Kite & Johnson, 1988; Kite, Stockdale, Whitley, & Johnson, 2005) as are physical abilities (Davis & Friedrich, 2010). For example, in memory taxing situations, older adults are perceived as being less credible and less accurate (Muller-Johnson, Toglia, Sweeney, & Ceci, 2007). The tendency to adjust our perceptions of capabilities of others based on appearance, whether unfounded or not, may influence another subjective perception: levels of trust placed in the individual's abilities. This pattern of prediction of capabilities is very relevant when users are presented with robots and unwittingly apply stereotypes; will they adjust their beliefs about its capabilities, and will it influence their trust? Accordingly, we might expect that when asked to judge the perceived capability of the robot, participants will be influenced by the age appearance. That is, consistent with aging stereotypes, capability ratings may be higher when the robot appears young compared to when the robot appeared old because cognitive and physical capability decline with age in healthy adults (Davis & Friedrich, 2010; Kite, Stockdale, Whitley, & Johnson, 2005). In addition, overall perceived capabilities of the robot will be higher when task performance is reliable. With regard to trust, trust in the robot will be highest when the task is stereotypically congruent with the robot's appearance (e.g. a younger robot performing a cognitive task instead of an older robot performing a cognitive task) and its performance was reliable. This is hypothesised because age appearance influences people's trust in automation (Pak, Fink, Price, Bass, & Sturre, 2012) and aging stereotypes are less likely to be activated while interacting with the younger robot.

1.4. Stereotypical thinking within Human-Robot interaction

There is evidence to suggest that stereotypical thinking applies within the HRI context. For example, individuals attribute stereotypical gender-specific traits to robots appearing to be male or female (Eyssel & Hegel, 2012; Tay, Jung, & Park, 2014; Kuchenbrandt, Häring, Eichberg, Eyssel, & André, 2014). Further, recent research suggests that implicit racial biases may be applied to robots in a manner consistent to which these biases are applied in human-human interaction (Bartneck et al., 2018). However, the application of aging stereotypes towards robots has been less studied and it is unclear if prior research is generalisable to this new technology context. Given that age is one of the most prominent characteristics noticed of a person (Fiske, 1998), the current study examined whether age stereotypes, induced by a robot's appearance, might cause stereotypic thinking.

Pak, McLaughlin, and Bass (2014) examined whether the physical appearance of an anthropomorphic software aid would activate stereotypic thinking and affect individuals' *trust* in a technological decision aid. They found that both younger and older adult participants trusted the older anthropomorphic aids more than the younger aids, the male aids more than the female aids, and more reliable applications than less reliable applications. Critically, stereotypic thinking, as measured by trust, was activated when perceptions of reliability of the aid were low or ambiguous. However, their study used a simple measure of stereotypic thinking (trust) rather than a multidimensional approach of the direct measurement of causal attributions and of perceived capabilities of the automated aid.

1.5. Remaining questions in aging and Human-Robot interaction

Studies investigating the psychological factors in human-robot interaction are still relatively new, therefore there are many gaps in the literature especially regarding the social influences on HRI. First, it is suggested that the physical embodiment component that is present in HRI can foster similar patterns of behavioural attribution compared to human-human interaction (e.g. attributions of intentionality, goal-setting; Froese & Ziemke, 2009; Ziemke, Thill, & Vernon, 2015; de Graaf, & Malle, 2017). Indeed, even after providing participants with higher levels transparency in terms of how a robot makes decisions, participants were still likely to attribute a robot behavioural outcome to that fact that it was 'thinking' (Wortham, Theodorou, & Bryson, 2017), suggesting that individuals make dispositional attributions even when situational information is available in HRI. However, these studies do not interpret behavioural and causal attributions through the lens of stereotype activation based on robot appearance. Next, studies in this area (e.g. Lee, 2003) typically have used measurements of stereotype activation that are not as robust compared to those used in the social cognition and psychological literature (e.g. Bieman-Copland, & Ryan, 1998). To our knowledge, none have used more rigorous measures of the application of stereotypes; that is, the effect of stereotypes on causal attributions, or the psychological processes and strategies that humans use to explain the actions of other people or entities (Kelley, 1973). Stereotypes powerfully influence the types of causal attributions that people make about the performance of others (Fiske & Taylor, 1991) and is a good indicator of whether one is applying a stereotype to a person or entity.

Second, although gender stereotypes have been well-studied using anthropomorphic technological aid paradigms, aging stereotypes have received much less attention (however, see Pak, McLaughlin, & Bass, 2014). Aging stereotypes are important because age is one of the most salient attributes noticed about a person (Fiske, 1998).

Third, studies of the perception of anthropomorphic technology have traditionally examined effects in college-aged samples. Previous research has shown age differences in susceptibility to the influence of anthropomorphic technology (Pak, Fink, Price, Bass, & Sturre, 2012). Although younger adult's levels of trust increased with anthropomorphic aids, older adults' trust levels did not differ between anthropomorphic automation and automation without human-like features. Novel studies with robots should compare younger and older adults' perceptions to verify whether stereotypes are present in each group. This is especially critical because much of future robot development is targeted towards the elderly (Broekens, Heerink, & Rosendal, 2009; Sharkey & Sharkey, 2012; Bemelmans, Gelderblom, Jonker, & De Witte, 2012).

1.6. The current study

Although there is evidence to suggest that stereotypes can affect perceptions and performance with anthropomorphised technological aids, we do not know how pre-existing *age* stereotypes will affect HRI. Next, it is unclear how trust is moderated by task type or domain in human-robot teaming. Although the automation literature affirms the important role of reliability on trust, to our knowledge there are very few studies explicitly investigating the moderating role of task type or domain on human perceptions of robots. Prior research has shown that task domain of automation has large effects on trust (Pak, Rovira, McLaughlin, & Baldwin, 2017). Finally, how does stereotyping technology affect perceptions of capabilities and the causal attributions made about performance?

The purpose of this study was to explore the extent to which younger and older adults applied age-based stereotypes to robots that appeared to be younger or older. The literature from social cognition and human factors are informative but there are still questions as to whether their results apply to the new domain of physical robots; specifically, whether the robot's appearance (age), task domain, and reliability of the robot's performance influence trust. In general, because this study is exploratory, we have only general hypotheses regarding the major effects. Namely, that the robot's appearance (age), level of reliability, and the task domain would affect trust towards a robot, the causal attributions that the individual makes about the robot's performance, and perceptions of the capability of the robot. To elaborate on these general expected effects:

1. We predicted that there would be a main effect of participant age, robot age, and reliability on causal attributions. The measurement of causal attributions (described further in Section 2.2.3) involves separately measuring the extent to which one makes a situational or dispositional attribution upon viewing an event. We expected that:
    a. consistent with the social cognition literature, older adults would make significantly higher dispositional attributions than younger adult participants,
    b. participants would make higher dispositional attributions when the robot appears older,
    c. and dispositional ratings would be higher for unreliable task performance than reliable task performance. This is because older adults are more likely to make dispositional (i.e. internal) attributions of blame when an outcome of an event is perceived as negative (the unreliable condition) and when their beliefs are violated (i.e. when an older looking robot performs the cognitive and physical tasks; Blanchard-Fields, Hertzog, & Horhota, 2012).

2. Perceived capability of the robot was expected to depend on the robot's age appearance. That is, consistent with aging stereotypes, capability ratings were expected to be higher when the robot appeared young compared to when the robot appeared old because cognitive and physical capability decline with age in healthy adults (Davis & Friedrich, 2010; Kite, Stockdale, Whitley, & Johnson, 2005). We also hypothesised that perceived capabilities would be higher when task performance is reliable.

3. Trust in the robot would be highest when the task was stereotypically congruent with the robot's appearance (e.g. a younger robot performing a cognitive task instead of an older robot performing a cognitive task) and its performance was reliable. This was hypothesised because age appearance influences people's trust in automation (Pak, Fink, Price, Bass, & Sturre, 2012) and aging stereotypes are less likely to be activated while interacting with the younger robot.
4. Task domain was treated as an exploratory variable. However, based on automation trust literature suggesting that trust in robot capability might depend on the domain in which they are placed (e.g. industry, entertainment, social; Schaefer, Sanders, Yordon, Billings, & Hancock, 2012; Pak, Rovira, McLaughlin, & Baldwin, 2017). To limit the complexity of the study, we only included two broad domains of tasks (physical, cognitive). We hypothesised that there would be a main effect of task domain such that participants would have more trust in the robot and have higher ratings of perceived capability when the robot performs physical tasks.

2. Method

2.1. Participants

Sixty younger adults ages 18–22 ($M = 18.65$, $SD = 1.01$) and 43 older adults ages 65–79 ($M = 70.53$, $SD = 3.96$) were recruited for this study. Younger adults were undergraduate college students who received extra credit for participation. Older participants were normatively aging older adults recruited from the community and received $15 for their participation. Clemson University's Institutional Review Board approved the experiment.

Eleven younger adults and seven older adults were eliminated from analysis due to missing data. The remaining 49 younger adults and 36 older adults were included in data analysis. The mean age of the younger group was 18.7 ($SD = 1.05$) and the older group was 70.8 ($SD = 4.03$). Descriptive statistics of participant characteristics are shown in Table 1.

Table 1. Participant characteristics by age group and gender.

	Younger adults ($n = 49$)				Older adults ($n = 36$)			
	Female ($n = 39$)		Male ($n = 10$)		Female ($n = 22$)		Male ($n = 14$)	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Age	18.44	0.79	19.6	0.43	69.86	0.8	72.14	0.13
CPRS*[a]	51.54	0.71	52.5	0.78	49.62	0.04	51.33	0.08

Note: *No significant age or gender differences. [a]Scores could range from 16 indicating low complacency potential to 80 indicating high complacency potential (Singh, Molloy, & Parasuraman, 1993).

2.2. Measures

2.2.1. Individual differences in attitudes towards automation (CPRS)

To describe participants' pre-existing attitudes towards automated systems, we used the Complacency Potential Rating Scale (CPRS; Singh, Molloy, & Parasuraman, 1993). CPRS is a 16-item scale ($\alpha = 0.87$) that measures complacency towards common types of automation. Participants responded to the extent they agreed with statements about automation on a scale of 1–5. The CPRS score was a sum of the responses where higher values indicated higher complacency potential. There were no significant differences between gender or age groups on the CPRS measure.

2.2.2. Causal attribution measurement

Causal attributions were measured using a paradigm adapted from Blanchard-Fields, Chen, Schocke, and Hertzog (1998). Participants were asked to indicate the degree to which either dispositional factors (person or actor-related) of the characters or situational factors (non-actor related) influenced the outcome of the scenario. The measure contained 6 items: 3 items measuring dispositional attributions ($\alpha = 0.90$) and 3 items measuring situational attributions ($\alpha = 0.80$). Specifically, participants indicated the extent to which: (a) the robot was responsible for the final outcome, (b) the robot was to blame for the final outcome, (c) the final outcome was due to personal characteristics of the robot, (d) the final outcome was due to characters in the story other than the robot, (e) the final outcome was due to something other than the characters in the story, and (f) both the personal characteristics of the robot and something other than the robot contributed to the final outcome. Participants responded using a Likert scale from 1 (very little) to 7 (very much). The averaged responses from a to c, represented dispositional attributions of performance while averaged responses to d–f, represented situational attributions of the final outcome. The higher the score on these two aspects, the higher the degree of either dispositional attributions or situational attributions.

2.2.3. Measurement of perceived capabilities

Perceived capabilities of the robot were measured using a list of 10 items ($\alpha = 0.91$) that spanned potential capabilities. Participants were asked, 'Based on the robot's behavior in the video you just watched, what other activities could the robot complete'? Participants were asked specifically the likelihood that the robot could carry out a variety of cognitive and physical tasks. For example, participants were asked, 'Based on the robot's performance, could it also recommend stock investment picks'? (a cognitive task) or 'Based on the robot's performance, could it also vacuum a room' (a physical task). Participants rated the extent to which they thought the robot could perform these tasks on a 1–7 Likert scale ranging from 'Definitely No' to 'Definitely Yes' with higher scores indicating increased perceptions of capabilities.

2.2.4. Trust measurement

Trust was measured using a single question modelled after Lee and Moray (1994) asking participants how much they trusted the robot portrayed in the scenario. Responses were recorded on a Likert scale from 1 (not at all) to 7 (very much). The higher the participants' ratings, the more their subjective trust in the robot.

2.2.5. Robot scenarios

Using a method commonly used in the human-human social cognition and aging literature (Chen & Blanchard-Fields, 1997; Follett & Hess, 2002; Ruffman, Murray, Halberstadt, & Vater, 2012), video scenarios were used to assess participants' attitudes towards the robot's behaviour and appearance. To assess perceptions of attribution, estimates of capabilities, and trust, all subjective assessments, we used scenario-based methodology commonly used in the sociological literature when the desire is to assess how independent factors might affect perceptions (Auspurg & Hinz, 2014). In this technique, independent variables (i.e. factors or dimensions) are treated as statistically independent, making it possible to identify and separate their influences on judgments (Rossi & Anderson, 1982). After presenting a scenario or scenario that illustrates the outcome of a robot collaboratively completing a task with a human, subjective perceptions about the scenario are measured. This method is well suited for the measurement of subjective constructs (e.g. trust, attributions) that are influenced by multiple, interacting factors. In addition to sociological research, variations of this method have also been used in human factors research (Endsley & Kiris, 1995).

In exploring the use of scenarios in human-robot research, Xu et al. (2015) suggested several guidelines. Notably, they suggest that scenarios that assess human attitudes of the robot should be of high fidelity and capture the uncertainties of a live interaction by recording a real human-robot interaction. Our scenarios captured a real human-robot interaction but presented them in sequential still images. In addition, Xu et al. (2015) state that if social aspects of the robot are of importance, interactivity (i.e. research participants physically interacting with robot) is not a strong modulating factor. A slideshow presentation was selected for both practical and theoretical reasons. First, the Baxter robot must undergo significant programming in order to perform the simplest of tasks, such as gripping a block at a specific location on a flat surface. Therefore, programming the robot to complete full circuit tasks would have required extensive time. Theoretically, our purpose was to apply a well-researched area, social cognition and aging stereotypes, to a novel field, HRI. Therefore, we tried to replicate experimental paradigms that require situational ambiguity within the stimuli. The slideshow format provided a means to present sequences of the robot's behaviour while still allowing for ambiguity.

In the current study, a human and robot were shown to be collaborating on a task in each scenario. Each scenario included manipulations of the age of the robot (younger, older), the domain of the collaborative task (cognitive, physical) with two tasks per domain, and the reliability of the robot's performance (low, high). To ensure that tasks were clearly perceived as cognitive or physical, and that the tasks were perceived to be of the same general level of difficulty, we pilot-tested the tasks to roughly equate the relative level of perceived difficulty before running the study. The pilot test consisted of asking four younger adults (college-aged) and three older adults (over age 65) whether a list of tasks was mostly cognitive (1) or physical (7) on a Likert scale. Some example tasks were 'creating a meal plan', 'mowing a lawn', 'threading a needle', or 'playing chess'. From this list of tasks, we selected tasks that were rated about two standard deviations above (very physical tasks) and below (very cognitive tasks) the mean rating of 3. These tasks were then subjected to another pilot study where 4 participants (college-aged) were asked to describe the difficulty level of the task on a likert scale. We then selected the tasks that were about the mean level (3) on this difficulty scale. The resulting tasks are illustrated in Table 2 along with how they were classified. Table 2 describes the scenarios (each participant saw all 16 scenarios). We indirectly measured the level of human stereotype activation by examining three subjective opinions of the robot: causal attributions regarding the robot's performance, the perceived capability of the robot, and the level of trust participants exhibited towards the robot.

Table 2. Description of all 16 scenarios.

Cognitive task		Physical task	
Low reliability	High reliability	Low reliability	High reliability
Young robot video scenarios			
Recycling (sorting recyclables). Robot mis-sorts.	Recycling (sorting recyclables). Robot properly sorts.	Stacking boxes (stacking boxes from one location to another). Robot drops the box.	Stacking boxes (stacking boxes from one location to another). Robot successfully stacks boxes.
Laundry (separating white and coloured clothing). Robot mis-sorts.	Laundry (separating white and coloured clothing). Robot properly sorts.	Changing light bulb. Robot breaks the lightbulb.	Changing light bulb. Robot successfully changes the lightbulb.
Older robot video scenarios			
Recycling (sorting recyclables). Robot mis-sorts.	Recycling (sorting recyclables). Robot properly sorts.	Stacking boxes (stacking boxes from one location to another). Robot drops the box.	Stacking boxes (stacking boxes from one location to another). Robot successfully stacks boxes.
Laundry (separating white and coloured clothing). Robot mis-sorts.	Laundry (separating white and coloured clothing). Robot properly sorts.	Changing light bulb. Robot breaks the lightbulb.	Changing light bulb. Robot successfully changes the lightbulb.

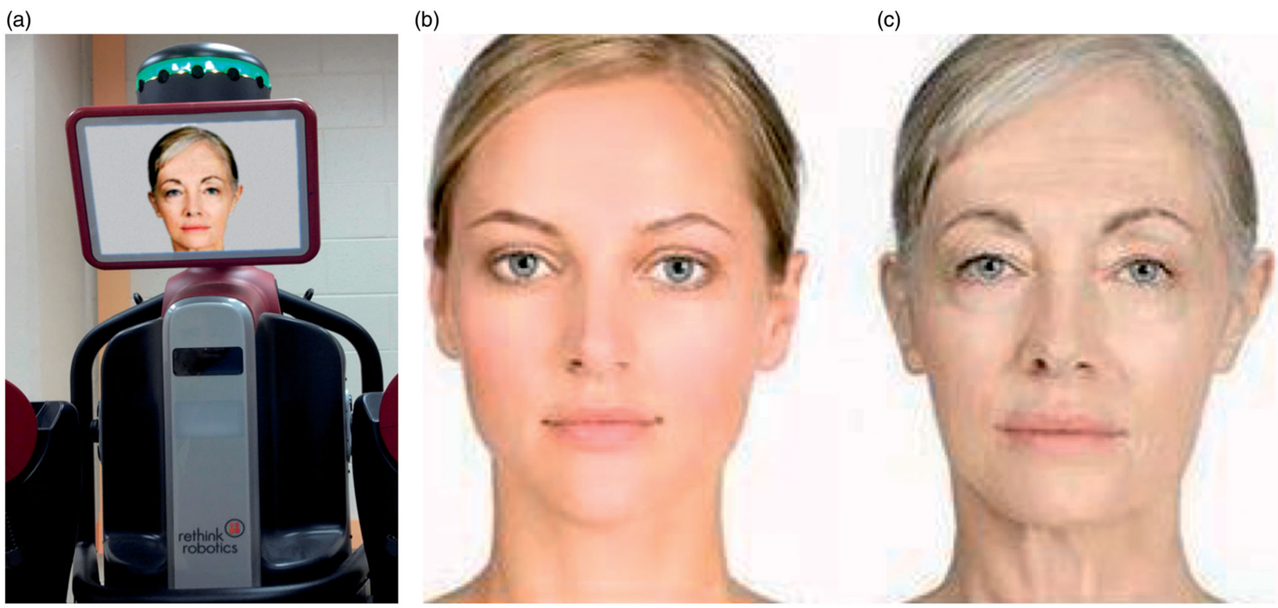

Figure 1. (a) (left). Baxter robot representing an older adult. (b) (middle). Facial stimuli used for younger adult. (c) (right). Facial stimuli used for older adult.

The robot used in this study was the Baxter robot manufactured by Rethink Robotics. Baxter is a manufacturing robot that can complete tasks that involve assembly and object organisation. Age of the robot was manipulated by digitally super-imposing one of two faces onto the display of the robot (Figure 1(a)). The younger and older adult faces used in the study are shown in Figure 1(b,c). Because the current study did not manipulate the gender of the robot, the facial stimuli for both the younger and older condition were female. In order to control for potential confounds for different faces, the faces selected for this study represented an age progression of the same female.

Each scenario contained a slideshow of pictures, with a fading transition between slides, portraying a human and a robot completing a collaborative task. Each video scenario was about 1 minute in length. The opening scenes included a wide shot, introducing the positioning of the human and robot as well as the collaborative task. In order to avoid any age or gender biases of the human actor, only the actor's arms and hands were shown while aiding in the collaborative task. The next shot included a close up of the robot's trunk, arms, and face. Finally, the human and the robot engage in the task. The final shot of the slideshow included information about whether the task was performed reliably. If the task was performed reliably, the final shot showed the task successfully completed. If the task was not performed reliably, the final shot showed the final outcome being incorrectly completed or unfinished. For example, in the light bulb

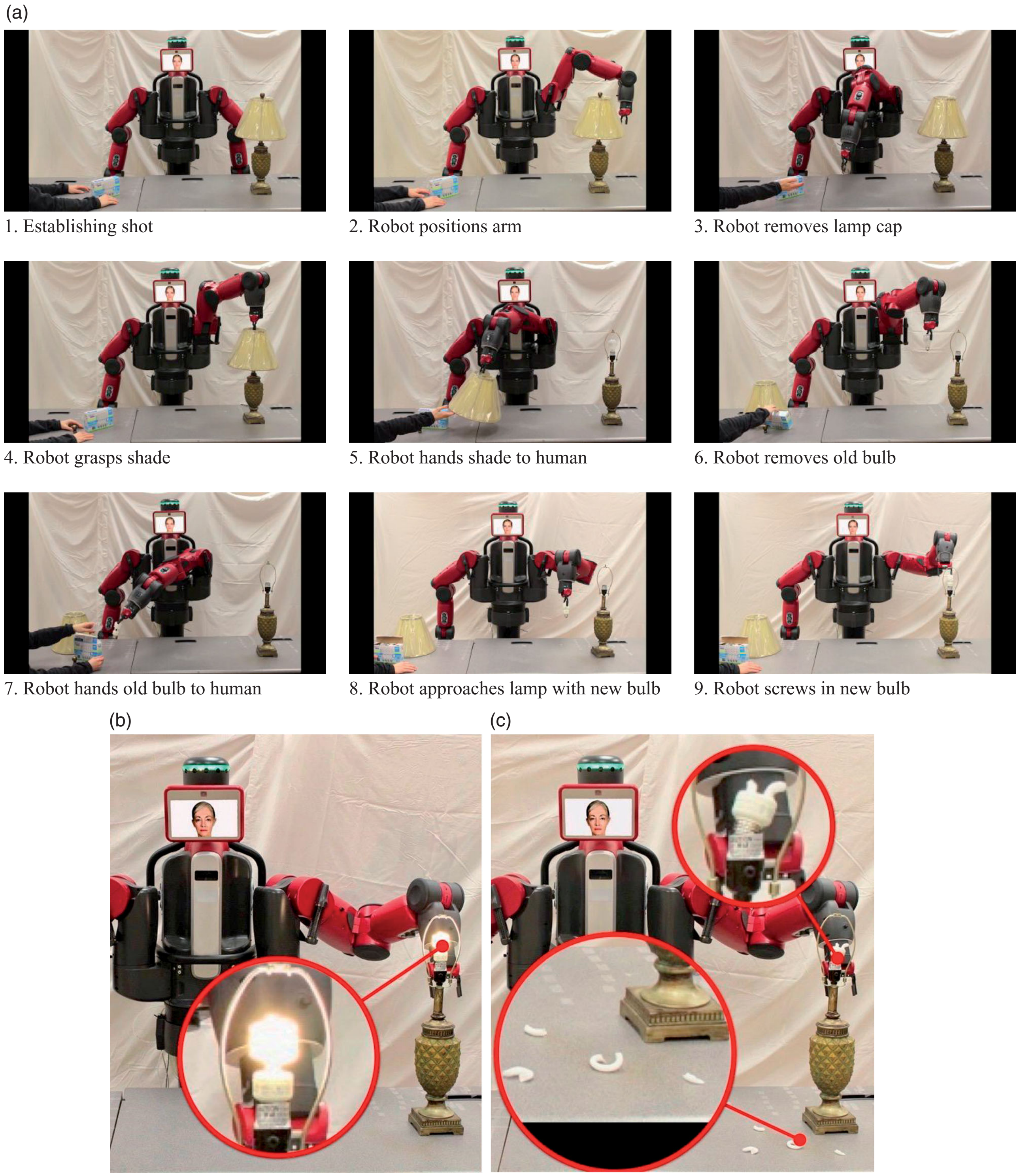

Figure 2. (a) Still keyframes from light bulb changing scenario (younger robot illustrated). (b) Light bulb changing task completed successfully; note the illuminated bulb (older robot illustrated). The inset zoom is for manuscript purposes and was not shown to participants. (c) Light bulb changing task not successful; note broken lightbulb on the table surface (older robot illustrated). The inset zoom is for manuscript purposes and was not shown to participants.

changing scenario (Figure 2(a)), reliable performance was portrayed with a photograph showing an illuminated, properly installed light bulb in the lamp (Figure 2(b)). In the unreliable scenario, the final photograph showed the light bulb broken into pieces on the table (Figure 2(c)).

During the study, each video scenario was presented in the centre of the screen and was not user-controllable. After participants viewed a video scenario, the questions and rating scales appeared in the lower half of the screen. This procedure was repeated for all 16 scenarios which were presented in a random, counterbalanced order. The study was administered using Qualtrics web platform.

2.2.6. Manipulation checks

Following the presentation of each robot scenario, participants responded to a series of manipulation checks regarding the appearance of the robot as well as the task performed. Specifically, participants were asked whether the robot appeared to be younger or older on a 7-point Likert scale. Participants were then asked whether the robot reliably performed the task to completion using a 7-point Likert scale ranging from 'No' to 'Yes'.

2.3. Design and procedure

The study was a 2 (participant age: younger, older) × 2 (robot age: young, old) × 2 (task domain: cognitive, physical) × 2 (robot reliability: low, high) mixed-model design, with participant age as the between-subjects variable. The within-subjects factors were manipulated in the study. The task domain dimension had two levels: cognitive and physical. These levels were selected in order to encompass the range of task domains within the HRI literature. Within those two domains, participants viewed the robots completing two separate tasks. That is, the robots completed two different cognitive tasks and two different physical tasks throughout the study. The two physical tasks included moving boxes from one location to another and changing a light bulb. The two cognitive tasks included sorting recycling and separating laundry.

Following consent, the experimenter e-mailed participants a personalised web link in order for them to complete a unique version of the study. The study was completed in their home so no lab visit was necessary. Participants worked through the study at their own pace. However, they were instructed to complete the study in one sitting. During the study, participants viewed randomly presented scenarios and answered each question after the completion of the slideshow. Participants completed the CPRS at the conclusion of the study.

3. Results and discussion

The following analyses are organised by the three main dependent variables of interest: causal attributions of robot performance, capability estimations, and trust. A commonly encountered problem with the use of analysis of variance procedures (ANOVA) in studies with many independent variables is a large number of statistically significant findings, making interpretation difficult. Simple measures of statistical significance such as p-value criterion do not indicate the size of the effect, which is arguably more meaningful. To more easily examine what was *practically* significant (i.e. because of greater effect size and thus more likely to appear in the real world) in our complex design, we explored all statistically significant effects ($p < 0.001$) but that also had a medium effect size as indicated by partial eta squared. Partial eta squared (η_p^2) indicates the amount of unique variance accounted for in the dependent variable by the independent variable after controlling for the influence of other variables and is equivalent to R^2; a higher value indicates a larger effect of the independent variable on the dependent variable.

While published guidelines indicate that a threshold of $R^2 = 0.14$, or 14% unique variance accounted for, is considered a *large* effect size (Cohen, 1988; Richardson, 2011), we opted to use a medium-effect size criterion. This decision was based on a meta-analysis of published studies examining trust with robots (Hancock et al., 2011) which found a *medium* effect size of $r = 0.26$ (equivalent to $R = 0.06$), which corresponds to 6% variance in trust due to global factors such as reliability and appearance. Thus, aside from any a priori hypotheses (which we will examine using $p < 0.05$), our exploratory analysis will only discuss significant effects (using a more stringent $p < 0.001$) that also have a partial eta squared value of 0.06 and above. This strategy makes the results easier to understand since we will not discuss results that may be statistically significant ($p < 0.001$) but account for little variance in the dependent variable (i.e. have an effect size of less than 0.06). This approach reduces the likelihood of type I error (false positives) but increases the likelihood of type II error (false negatives) but we felt this was a worthwhile trade-off given the complexity of the design and sheer volume of results that result from the interactions of variables.

For each analysis, a summary ANOVA table, with *p*-values, and effect sizes, is presented to get a clearer understanding of the practically significant effects. In discussing the results, unless we had a priori hypotheses we will not discuss main effects or any lower order interactions as they are not interpretable in the presence of a significant higher-order interaction.

Table 3. ANOVA summary table of dispositional ratings.

Sources	SS	df	MS	F	p	η_p^2
Between subjects						
User age	**191.22**	**1**	**191.22**	**5.92**	**0.02**	**0.07**
S within-group error	2680.64	83	32.30			
Within subjects						
Robot age	**5.42**	**1**	**5.42**	**24.92**	**0.00**	**0.23**
Reliability	**28.81**	**1**	**28.81**	**38.67**	**0.00**	**0.32**
Domain	**88.38**	**1**	**88.38**	**124.14**	**0.00**	**0.60**
Robot age × User age	0.01	1	0.01	0.02	0.88	0.00
Robot age × Reliability	0.16	1	0.16	0.63	0.43	0.01
Robot age × Domain	**3.84**	**1**	**3.84**	**12.80**	**0.00**	**0.13**
Reliability × User age	1.58	1	1.58	2.13	0.15	0.03
Reliability × Domain	**17.40**	**1**	**17.40**	**39.51**	**0.00**	**0.32**
Domain × User age	2.45	1	2.45	3.45	0.07	0.04
Robot age × Reliability × User age	0.10	1	0.10	0.41	0.52	0.01
Robot age × Domain × User age	0.32	1	0.32	1.07	0.30	0.01
Robot age × Reliability × Domain	**2.60**	**1**	**2.60**	**10.24**	**0.00**	**0.11**
Reliability × Domain × User age	0.04	1	0.04	0.09	0.76	0.00
Robot age × Reliability × Domain × User age	0.14	1	0.14	0.55	0.46	0.01
Error (Robot age)	18.05	83	0.22			
Error (Reliability)	61.84	83	0.75			
Error (Domain)	59.10	83	0.71			
Error (Robot age × Domain)	24.86	83	0.30			
Error (Robot age × Reliability)	20.44	83	0.25			
Error (Reliability × Domain)	36.55	83	0.44			
Error (Robot age × Reliability × Domain)	21.03	83	0.25			

Note. Bolding indicates practically significant effects ($p < 0.001$ and an effect size, η_p^2, greater than 0.06). *S*: subjects.

All post-hoc analyses for significant effects were adjusted using Bonferroni corrections.

We first analysed whether our primary manipulations had their intended effects. Manipulation checks for perceived age of the robot (t (84) = 14.29, $p < 0.001$), perceived reliability of the robot (t (84) = 29.56, $p < 0.001$), and perceived task domain of the robot (t (84) = 7.49, $p < 0.001$) verified that the manipulations had their intended effects in the expected directions. In the analyses that follow, we first examined how causal attributions differed as a function of experimental manipulations (robot age, reliability, task type) and person-factors (age group). Then we subsequently examine the other dependent measures (capability estimation, trust).

3.1. Causal attributions

To investigate how the appearance and actions of the robot would influence how people attribute performance to the robot (dispositional or situational), the two measurements of dispositional and situational attributions are customarily separated and treated as different dependent variables (see Blanchard-Fields, 1994; Blanchard-Fields, Chen, Schocke, & Hertzog, 1998). A 2 (participant age: younger, older) × 2 (robot age: young, old) × 2 (robot reliability: low, high) × 2 (task domain: cognitive, physical) mixed repeated measures ANOVA was performed on a mean of the three items representing dispositional attributions, and a separate 2 (participant age: younger, older) × 2 (robot age: young, old) × 2 (robot reliability: low, high) × 2 (task domain: cognitive, physical) mixed repeated measures ANOVA was conducted on a mean of the three situational attribution items.

3.1.1. Dispositional attributions

Dispositional ratings indicate the likelihood of attributing robot task performance to the robot rather than the situation. The repeated measures ANOVA (Table 3) for dispositional attributions revealed a main effect of participant age group (F (1, 83) = 5.921, $p < 0.017$, $\eta_p^2 = 0.067$), indicating that, contrary to our hypotheses, younger adults ($M = 6.02$, $SD = 2.65$) made significantly higher dispositional attributions than older adults ($M = 4.95$, $SD = 3.09$). In addition, the other main independent variables were significant (robot age, reliability, task domain). However, while there were these significant main effects and lower order interactions (e.g. the 2-way interaction between robot age and domain), they are qualified by the presence of the significant 3-way interaction between robot age, reliability, and task domain (F (1, 83) = 10.24, $p = 0.002$, $\eta_p^2 = 0.110$).

The source of the 3-way interaction, illustrated in Figure 3(a,b), was a significant 2-way interaction between age of robot and reliability within the cognitive task domain (F (1, 83) = 39.513, $p < .001$, $\eta_p^2 = 0.323$) and within the physical task domain (F (1, 83) = 10.24, $p = 0.002$, $\eta_p^2 = 0.110$). For cognitive tasks, when task performance was reliable, participants made higher dispositional ratings when the robot

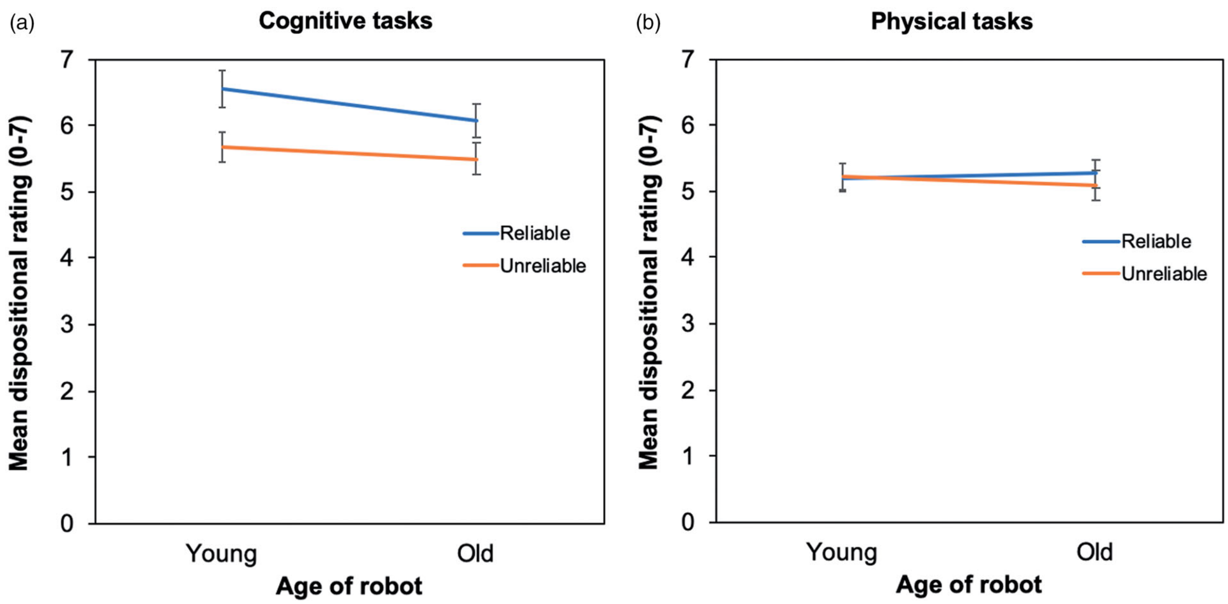

Figure 3. (a) Dispositional attribution as a function of robot age and reliability for cognitive tasks. (b) Dispositional attribution as a function of robot age and reliability for physical tasks.

appeared young ($M = 6.47$, $SD = 2.45$) compared to when the robot appeared older ($M = 5.98$, $SD = 2.33$). When performance was unreliable, however, participants made significantly less dispositional ratings for both younger ($M = 5.57$, $SD = 2.04$) and older appearing robots ($M = 5.39$, $SD = 2.20$). For physical tasks, there were no differences between high and low reliability when the robot appeared young ($p > .05$). When the robot appeared older, however, participants made significantly more dispositional attributions when the robot performed the task with high reliability ($M = 5.20$, $SD = 2.00$) compared to low reliability ($M = 5.02$, $SD = 2.01$).

While there is a general age effect on making dispositional attributions, the effect sizes show that the extent of making dispositional attributions was primarily driven by the appearance of the robot its perceived reliability, and the type of task. When the task was cognitive in nature, and the robot was reliable, participants made more dispositional attributions for younger robots compared to older robots; that is, participants attributed its performance to an internal characteristic. However, when the robot was older and performed reliably in the same cognitive task, they made a significantly lower dispositional attribution as compared to a younger robot. For physical tasks, when the robot appeared younger, there was no difference by robot reliability in dispositional attributions; that is, participants made equivalent dispositional attributions when the robot was successful or failed. However, when the robot appeared older, reliability was a key differentiator: participants made higher dispositional attributions when it was reliable than not.

3.1.2. Situational attributions

Situational attribution ratings indicate the likelihood of attributing robot task performance to the situation rather than inherent robot characteristics. The ANOVA for situational attributions (Table 4) revealed only significant main effects (bolded) and no interactions. First, there was a significant main effect of robot age such that participants made significantly more situational attributions about the robot's behaviour when the robot appeared younger ($M = 4.18$, $SD = 1.94$) than when the robot appeared older ($M = 4.01$, $SD = 1.79$). Next, the reliability of the robot affected situational attributions with reliable robots eliciting lower situational attributions ($M = 3.93$, $SD = 1.79$) than unreliable robots ($M = 4.16$, $SD = 1.95$). That is, when the robots were reliable, all participants tended to discount the situational factors (i.e. ease of task) but when it failed, participants tended to attribute failure to situational factors (i.e. difficult of the task). Finally, the significant effect of domain shows that cognitive tasks elicited more situational attributions ($M = 4.21$, $SD = 2.11$) than physical tasks ($M = 3.89$, $SD = 1.64$).

3.2. Capability estimation

Responses from the capabilities rating scales were summed within each condition to provide a total score of the robot's perceived capabilities.

Table 4. ANOVA summary table of situational ratings.

Sources	SS	*df*	MS	*F*	*p*	η_p^2
Between subjects						
User age	58.00	1	58.00	2.17	0.14	0.25
S within-group error	2218.19	83	26.73			
Within subjects						
Robot age	**4.32**	**1**	**4.32**	**10.90**	**0.00**	**0.12**
Reliability	**8.89**	**1**	**8.89**	**9.55**	**0.00**	**0.10**
Domain	**16.94**	**1**	**16.94**	**11.49**	**0.00**	**0.12**
Robot age × User age	0.55	1	0.55	1.39	0.24	0.02
Robot age × Reliability	0.01	1	0.01	0.03	0.86	0.00
Robot age × Domain	0.82	1	0.82	1.62	0.21	0.02
Reliability × User age	0.01	1	0.01	0.01	0.91	0.00
Reliability × Domain	3.67	1	3.67	4.10	0.05	0.05
Domain × User age	0.00	1	0.00	0.00	0.97	0.00
Robot age × Reliability × User age	0.69	1	0.69	1.92	0.17	0.02
Robot age × Domain × User age	0.06	1	0.06	0.12	0.73	0.00
Robot age × Reliability × Domain	0.07	1	0.07	0.16	0.69	0.00
Reliability × Domain × User age	1.23	1	1.23	1.37	0.25	0.02
Robot age × Reliability × Domain × User age	0.34	1	0.34	0.72	0.40	0.01
Error (Robot age)	32.93	83	0.40			
Error (Reliability)	77.25	83	0.93			
Error (Domain)	122.37	83	1.47			
Error (Robot age × Domain)	42.13	83	0.51			
Error (Robot age × Reliability)	29.82	83	0.36			
Error (Reliability × Domain)	74.39	83	0.90			
Error (Robot age × Reliability × Domain)	38.79	83	0.47			

Note. Bolding indicates practically significant effects ($p < 0.001$ and an effect size greater than 0.06). *S*: subjects.

Table 5. ANOVA summary table of capability estimations.

Sources	SS	*df*	MS	*F*	*p*	η_p^2
Between subjects						
User age	78.34	1	78.3	0.1	0.8	0.0
S within-group error	101314.56	83	1220.7			
Within subjects						
Robot age	**135.78**	**1**	**135.78**	**9.76**	**0.00**	**0.11**
Reliability	**7471.03**	**1**	**7471.03**	**34.42**	**0.00**	**0.29**
Domain	**229.45**	**1**	**229.45**	**11.66**	**0.00**	**0.12**
Robot age × User age	19.84	1	19.84	1.43	0.24	0.02
Robot age × Reliability	6.19	1	6.19	0.45	0.51	0.01
Robot age × Domain	**113.58**	**1**	**113.58**	**11.15**	**0.00**	**0.12**
Reliability × User age	39.26	1	39.26	0.18	0.67	0.00
Reliability × Domain	64.33	1	64.33	1.42	0.24	0.02
Domain × User age	19.45	1	19.45	0.99	0.32	0.01
Robot age × Reliability × User age	26.75	1	26.75	1.94	0.17	0.02
Robot age × Domain × User age	31.68	1	31.68	3.11	0.08	0.04
Robot age × Reliability × Domain	94.03	1	94.03	2.08	0.15	0.02
Reliability × Domain × User age	30.33	1	30.33	2.39	0.13	0.03
Robot age × Reliability × Domain × User age	4.79	1	4.79	0.38	0.54	0.01
Error (Robot age)	1154.38	83	13.91			
Error (Reliability)	18016.79	83	217.07			
Error (Domain)	1634.06	83	19.69			
Error (Robot age × Domain)	845.65	83	10.19			
Error (Robot age × Reliability)	1143.60	83	13.78			
Error (Reliability × Domain)	3753.44	83	45.22			
Error (Robot age × Reliability × Domain)	1055.03	83	12.71			

Note. Bolding indicates practically significant effects ($p < 0.001$ and an effect size greater than 0.06). *S*: subjects.

A 2 (participant age: younger, older) × 2 (robot age: young, old) × 2 (robot reliability: low, high) × 2 (task domain: cognitive, physical) mixed repeated measures ANOVA (summarized in Table 5) revealed a significant main effect of reliability (F (1, 83) = 34.418, $p < 0.001$, $\eta_p^2 = 0.293$). In accordance with our hypothesis, participants rated robots that performed a task reliably ($M = 32.07$, $SD = 14.93$) as having higher capabilities than those that performed a task unreliably ($M = 25.36$, $SD = 12.05$). Additionally, there was a significant interaction between robot age and task domain (F (1, 83) = 11.147, $p = 0.001$, $\eta_p^2 = 0.118$) on perception of capabilities. The source of this interaction, illustrated in Figure 4, was that when the robot appeared young, the robot carrying out cognitive tasks was perceived as having more capabilities ($M = 30.17$, $SD = 13.41$) than when carrying out physical tasks ($M = 28.16$, $SD = 12.26$). However, when the robot appeared older, there were no differences in capability ratings between cognitive ($M = 28.43$,

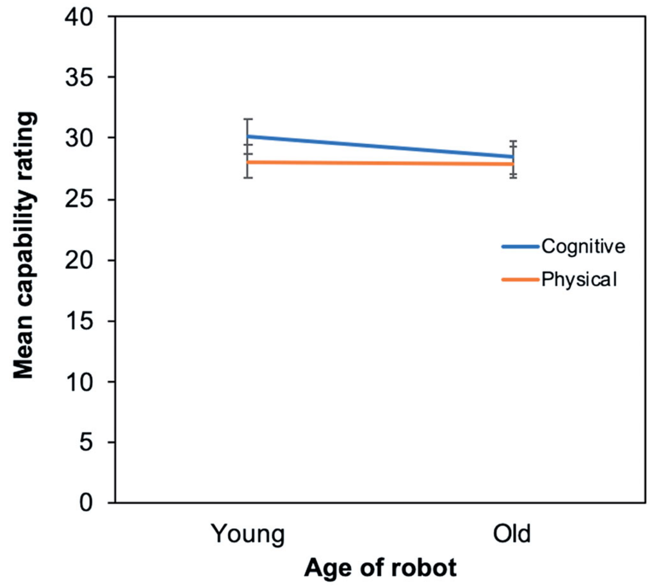

Figure 4. Age of robot × domain.

$SD = 12.81$) and physical task domains ($M = 28.09$, $SD = 12.37$). Perceptions of capabilities for cognitive tasks were also significantly higher when the robot appeared younger than when the robot appeared older.

3.3. Trust

The trust data was subjected to a 2 (participant age: younger, older) × 2 (robot age: young, old) × 2 (robot reliability: low, high) × 2 (task domain: cognitive, physical) mixed repeated measures ANOVA. A summary of the results of the ANOVA are shown in Table 6. The highest-order, practically significant interaction was between reliability and domain, ($F(1, 83) = 24.30$, $p < 0.001$, $\eta_p^2 = 0.22$). The interaction is illustrated in Figure 5. Pairwise comparisons, adjusted for multiple comparisons, showed that the source of the interaction was that when the robot was unreliable, there was a significant difference in trust, $F(1, 83) = 24.39$, $p < 0.001$, $\eta_p^2 = 0.23$, by task domain with less trust in cognitive ($M = 2.14$) tasks compared to physical tasks ($M = 2.65$). However, when the robot was reliable, there was no significant difference in trust as a function of task domain.

4. Conclusion

This study examined how pre-existing age stereotypes affected older and younger adults' perceptions of robots. Previous research has shown that stereotypes can affect performance and interactions with anthropomorphised technological aids. This study attempted to extend these findings to the HRI domain. It was hypothesised that trust in the robot would be highest when the task was stereotypically congruent with the robot's appearance and its performance was reliable. In summary, we found mixed support of our hypotheses which showed that human perceptions of robots is more complex, and may not be as simple as direct human application of existing stereotypes.

Predictions of causal attributions were based on previous social cognition literature (Blanchard-Fields, 1996). Therefore, we expected that dispositional attributions would be highest for older adult participants when the robot appeared older and performed tasks unreliably. However, this hypothesis was not supported. Contrary to our predictions, younger adults made significantly higher dispositional attributions than older adults. Overall, dispositional ratings were highest when a young-appearing robot reliably performed a cognitive task.

Because people attribute human-like qualities to technology, it is often the case that social constructs such as trust or stereotyping affect human-automation interaction similarly to the ways in which they affect human-human interaction. However, our findings suggests a major caveat in applying human theories of social cognition to technology. Specifically, individuals are more likely to 'give credit' to a robot for reliable performance as opposed to blaming it for unreliable performance. This represents one of possibly many ways that human-robot perceptions differ from human-human. However, consistent with widely-held human age stereotypes, participants were most likely to give credit to the robot when it appeared young and reliably performed a cognitive task.

Table 6. ANOVA summary table of trust.

Sources	SS	df	MS	F	p	η_p^2
Between subjects						
User age	0.29	1	0.29	0.23	0.88	0.00
S within-group error	1066.15	83	12.85			
Within subjects						
Robot age	0.09	1	0.09	0.21	0.65	0.00
Reliability	**1145.80**	**1**	**1145.80**	**202.50**	**0.00**	**0.71**
Domain	2.98	1	2.98	4.47	0.04	0.05
Robot age × User age	1.64	1	1.64	3.82	0.05	0.04
Robot age × Reliability	0.14	1	0.14	0.47	0.49	0.01
Robot age × Domain	0.09	1	0.09	0.28	0.60	0.00
Reliability × User age	4.13	1	4.13	0.73	0.40	0.01
Reliability × Domain	**24.30**	**1**	**24.30**	**23.34**	**0.00**	**0.22**
Domain × User age	1.29	1	1.29	1.93	0.17	0.02
Robot age × Reliability × User age	0.00	1	0.00	0.00	0.97	0.00
Robot age × Domain × User age	1.99	1	1.99	6.64	0.01	0.07
Robot age × Reliability × Domain	0.01	1	0.01	0.02	0.88	0.00
Reliability × Domain × User age	6.24	1	6.24	5.99	0.02	0.07
Robot age × Reliability × Domain × User age	0.01	1	0.01	0.02	0.88	0.00
Error (Robot age)	35.67	83	0.43			
Error (Reliability)	469.63	83	5.66			
Error (Domain)	55.43	83	0.67			
Error (Robot age × Domain)	24.82	83	0.30			
Error (Robot age × Reliability)	24.79	83	0.30			
Error (Reliability × Domain)	86.40	83	1.04			
Error (Robot age × Reliability × Domain)	35.67	83	0.43			

Note. Bolding indicates practically significant effects ($p < 0.001$ and an effect size greater than 0.06). *S*: subjects.

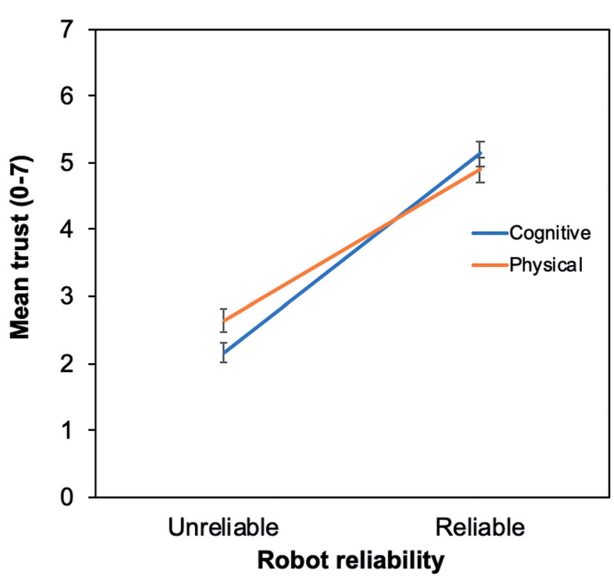

Figure 5. Two-way interaction between robot reliability and task domain on trust.

Situational attribution ratings indicate the likelihood of attributing robot task performance to the situation rather than inherent robot characteristics. It is important to note that dispositional and situational causal attributions are *not* mutually exclusive. Optimal causal reasoning involves consideration of both the dispositional characteristics of the actor and the external, situational factors (Fiske, 1993; Grossmann & Ross, 2018). Situational attributions followed a similar pattern as dispositional attributions such that participants made more situational attributions when the robot appeared young. Participants were also more likely to attribute task performance to situational factors when a cognitive task was performed reliably (regardless of the age of the participant or robot). Therefore, our results suggest that situational factors also influence adults' perceptions of causal reasoning. The fact that our dispositional and situational attribution patterns are complementary suggests that adults are able to attribute cause in a multidimensional way that is considered to be more ideal and accurate (Fiske, 1993) within the HRI context.

It is well established that individuals are more likely to place overdue emphasis on dispositional factors when a situation is ambiguous (i.e. the relative contributions of the actor and the contributions of the situation on a final outcome are unclear; Blanchard-Fields, 1994; Trope, 1986). In our slideshow scenarios, although we presented photographs of the final outcome in each scenario, the stimuli were ambiguous regarding the human collaborator's (a situational factor) influence over the final outcome. We also did not include any internal information about the robot's programming or instructions. Therefore, we believe our stimuli were ambiguous enough to allow participants to place overdue influence on the robot's internal qualities in the predicted direction. However, because the results were contrary to our hypothesis, perhaps individuals attribute causal attributions differently within the HRI context. From a design perspective, robots that appear younger, and reliably perform cognitive tasks are more likely to yield more optimal attribution patterns that consider both the dispositional qualities of the robot as well as external influences of the situation.

Although task domain was treated as an exploratory variable in our study, the current finding is consistent with the literature that trust in adults' cognitive abilities tends to decrease with advancing age (Lineweaver & Hertzog, 1998). It is surprising, however, that the effect of aging stereotypes did not affect younger adults' trust ratings. The aging literature suggests that the presence of aging stereotypes is also predicted by level of contact with aging individuals (Hale, 1998; Levy, 2018). This idea could also relate to level of contact with automation. It is documented that younger adults are more likely to own and interact with technology (Zickuhr, 2011; Friemel, 2016) and in-home robots (Sung, Grinter, Christensen, & Guo, 2008). Therefore, younger adults' levels of trust might be more influenced by their level of contact and familiarity with technology in general rather than the appearance of the robot.

Participants trusted the robot significantly more when performance was reliable, partially supporting the first hypothesis. Again, however, this effect was moderated by participant age and task domain. Although younger adults' trust ratings were resistant to changes in task domain and reliability, older adults are affected by these changes. When reliability was low, older adults trusted robots that performed physical tasks more than cognitive tasks. Conversely, when reliability was high, older adults trusted robots that performed cognitive tasks significantly more than those that performed physical tasks. This suggests that although all participants' trust ratings are sensitive to reliability in the expected direction, older adults' trust in robots is sensitive to reliability as a function of task domain. This supports the idea that trust in automation depends on the domain in which it is placed (e.g. industry, entertainment, social; Schaefer, Sanders, Yordon, Billings, & Hancock, 2012; Pak, Rovira, McLaughlin, & Baldwin, 2017). These findings are interesting for a number of reasons. By applying aging stereotypes to robots, older adult participants may be attributing age-related qualities to the robot similarly to the way they would attribute these qualities to themselves or to their peers. In the aging stereotype literature, aging-related cognitive failures are perceived to indicate an inherent lack of ability that is difficult or impossible to mitigate (Bieman-Copland, & Ryan, 1998; Cuddy, Norton, & Fiske, 2005). Conversely, the extent of age-based stereotype threat within physical domains is unclear (Lamont, Swift, & Abrams, 2015), indicating that aging stereotypes are indeed multidimensional such that physical decline might not be perceived as negatively as a cognitive failure. This supports our finding that unreliable performance on a physical task is not catastrophic to older adults' trust in the robot.

It was expected that perceived capability ratings would be higher when the robot performed reliably and appeared young. Supporting our hypothesis, participants rated reliable performing robots as having higher capabilities than those that performed a task unreliably. Further, participants rated younger appearing robots as having more capabilities than older appearing robots. We also expected that participants would have higher ratings of perceived capabilities when the robot performed physical tasks. This hypothesis was not supported; instead, task type (domain) interacted with robot age. Robots had the highest amount of perceived capabilities when they appeared young and completed cognitive tasks. However, age stereotypes did influence capability ratings such that, compared to younger adult robots, perceived capabilities were significantly lower when the robot appeared older and performed cognitive tasks.

The general literature in aging and cognition suggest an enhanced (or at least preserved) wisdom with age compared to losses in cognitive abilities (for a review see Baltes & Staudinger, 1993). The distinction between losses in 'cognitive mechanics' balanced by gains in 'cognitive pragmatics' (Baltes & Staudinger) was initially described by Baltes and Baltes' (1990) theory of selection, optimisation, and compensation. Accordingly, it seems surprising that when our participants were presented with older appearing robots, they did not attribute greater wisdom to it in the form of greater perceived capabilities for cognitive tasks. This may be explained by the fact that while some aging and wisdom studies do show an older adult advantage (e.g. Blanchard-Fields et al., 2007), this advantage seems narrowly focussed on decision making that has an emotional or social component. This idea is supported by findings that show no enhanced wisdom stereotype for older physicians compared to younger (Shah & Ogden, 2006).

However, there are several methodological and analytical issues that temper whether we can conclude that our subjects truly applied pre-existing age stereotypes to robots. One limitation was due to not reaching our desired sample size. An a priori power analysis revealed that in order to achieve a moderate between-groups comparison effect size ($d = 0.5$), approximately 50 participants per group would be needed to reach statistical significance at the 0.05

level. After screening out participants, we only analysed data from 49 younger adults and 36 older adults. While our analysis method (ANOVA) is considered very robust to homogeneity of variance differences due to unequal sample sizes, it should temper our final conclusions.

Another limitation is that we did not assess pre-existing stereotypes held by our participants because a stereotype assessment could have biased participant ratings during the study. However, the social cognition literature consistently finds pervasive expectations of cognitive and physical decline with increasing age (Davis & Friedrich, 2010; Fiske, 2017). Related to this point, we also did not examine how participants perceived the age-progressed face. That is, did they perceive the robot as having chronologically aged, or did they perceive it as something else. For example, participants could have mis-perceived the aged robot face as more 'tired' compared to the young face, and not older.

Another limitation was our analysis strategy. Given the exploratory nature of this study and the resulting large number of effects, we chose to use an analysis strategy that focussed on the practically significant effects. However, this resulted in us leaving some statistically significant effects unexplained. A more conceptual criticism might be our decision to base our practical significance criterion by examining the literature in HRI, specifically our use of the medium effect size. The alternative was to base the criterion from the large body of human-human stereotype research. While this may be the optimal strategy if no HRI literature existed, given the presence of some social science research in HRI, we chose to focus on the results of that literature. In the end, our practical significance criterion ($p < 0.001$ and medium effect size) was very simple but one that could be considered a very stringent one for the social sciences.

Older adult participants were more susceptible to stereotypic thinking as measured by trust and perceived capabilities of the robot. From a design perspective, when it is important for users to maintain high levels of trust in imperfect automation, a younger appearing robot that performs more physical tasks would be optimal because it is less susceptible to large fluctuations in perceptions of trust as a function of stereotypic thinking. However, these findings are more applicable to older adult users who experienced fluctuations in trust as a function of reliability, appearance, and task domain.

Although young adults' trust ratings were not sensitive to the manipulations, stereotype research shows that people of all ages are susceptible to stereotypic thinking (Kite, Stockdale, Whitley, & Johnson, 2005; Berger, 2017). Therefore, a reasonable option would be to design to avoid activating age stereotypes, especially in the face of imperfect technology. Patterns of causal reasoning within the HRI context also differed from causal reasoning patterns found in human-human interaction. From an applied perspective, robots that appear young, and reliably perform cognitive tasks are more likely to yield more optimal attribution patterns of causal reasoning. In sum, these findings extend well-established findings regarding the application of age stereotypes to a novel domain, robotics, while suggesting a major caveat in applying human theories of social reasoning to technology. Future research may attempt to replicate these results with physical robots or other stereotypes.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This study was partially funded by the Air Force Office of Scientific Research (FA9550-12-1-0385) and the Clemson Human Factors Institute.

ORCID

Richard Pak http://orcid.org/0000-0001-9145-6991

References

Ashmore, R. D., and F. K. Del Boca. 1981. "Conceptual Approaches to Stereotypes and Stereotyping." In *Cognitive Processes in Stereotyping and Intergroup Behavior*. edited by D. L. Hamilton, 1–35. Hillsdale, NJ: Erlbaum.

Auspurg, K., and T. Hinz. 2014. Factorial Survey Experiments. Series Quantitative Applications in the Social Sciences No. 175.

Baltes, P. B., and M. M. Baltes. 1990. "Psychological Perspectives on Successful Aging: The Model of Selective Optimization with Compensation." *Successful Aging: Perspectives from the Behavioral Sciences* 1 (1): 1–34.

Baltes, P. B., and U. M. Staudinger. 1993. "The Search for a Psychology of Wisdom." *Current Directions in Psychological Science* 2 (3): 75–81.

Bargh, J. A. 1994. "The Four Horsemen of Automaticity: Awareness, Intention, Efficiency, and Control in Social Cognition." *Handbook of Social Cognition* 1: 1–40.

Bartneck, C., K. Yogeeswaran, Q. M. Ser, G. Woodward, R. Sparrow, S. Wang, and F. Eyssel. 2018. "Robots and Racism." In Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction (pp. 196–204). Chicago, IL: ACM. doi:10.1145/3171221.3171260.

Bemelmans, R., G. J. Gelderblom, P. Jonker, and L. De Witte. 2012. "Socially Assistive Robots in Elderly Care: A Systematic Review into Effects and Effectiveness." *Journal of the American Medical Directors Association* 13 (2): 114–120. doi:10.1016/j.jamda.2010.10.002.

Berger, R. 2017. "Aging in America: Ageism and General Attitudes toward Growing Old and the Elderly." *Open Journal of Social Sciences* 5 (8): 183–198. doi:10.4236/jss.2017.58015.

Bieman-Copland, S., and E. B. Ryan. 1998. "Age-Biased Interpretation of Memory Successes and Failures in Adulthood." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 53B (2): P105–P111. doi:10.1093/geronb/53B.2.P105.

Blanchard-Fields, F. 1994. "Age Differences in Causal Attributions from an Adult Developmental Perspective." *Journal of Gerontology 49* (2): P43–P51. doi:10.1093/geronj/49.2.P43.

Blanchard-Fields, F. 1996. "Causal Attributions across the Adult Life Span: The Influence of Social Schemas, Life Context, and Domain Specificity." *Applied Cognitive Psychology* 10 (7): 137–146. doi:10.1002/(SICI)1099-0720(199611)10:7<137::AID-ACP431>3.0.CO;2-Z.

Blanchard-Fields, F., Y. Chen, M. Schocke, and C. Hertzog. 1998. "Evidence for Content-Specificity of Causal Attributions across the Adult Life Span." *Aging, Neuropsychology, and Cognition* 5: 241–263. doi:10.1076/anec.5.4.241.770.

Blanchard-Fields, F., C. Hertzog, and M. Horhota. 2012. "Violate my Beliefs? Then You're to Blame! Belief Content as an Explanation for Causal Attribution Biases." *Psychology and Aging* 27 (2): 324–337. doi:10.1037/a0024423.

Blanchard-Fields, F., A. Mienaltowski, and R. B. Seay. 2007. "Age Differences in Everyday Problem-Solving Effectiveness: Older Adults Select more Effective Strategies for Interpersonal Problems." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 62 (1): P61–P64.

Brewer, M. B., and L. Lui. 1984. "Categorization of the Elderly by the Elderly." *Personality and Social Psychology Bulletin* 10 (4): 585–595. doi:10.1177/0146167284104012.

Broekens, J., M. Heerink, and H. Rosendal. 2009. "Assistive Social Robots in Elderly Care: A Review." *Gerontechnology* 8 (2): 94–103. doi:10.4017/gt.2009.08.02.002.00.

Campbell, A., J. E. Murray, L. Atkinson, and T. Ruffman. 2017. "Face Age and Eye Gaze Influence Older Adults' Emotion Recognition." *The Journals of Gerontology: Series B: Psychological Sciences and Social Sciences* 72 (4): 633–636.

Chen, Y., and F. Blanchard-Fields. 1997. "Age Differences in Stages of Attributional Processing." *Psychology and Aging* 12 (4): 694–703. doi:10.1037/0882-7974.12.4.694.

Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). New York: Academic Press.

Cuddy, A. C., M. I. Norton, and S. T. Fiske. 2005. "This Old Stereotype: The Pervasiveness and Persistence of the Elderly Stereotype." *Journal of Social Issues* 61 (2): 267–285. doi:10.1111/j.1540-4560.2005.00405.x.

Czaja, S. J., N. Charness, A. D. Fisk, C. Hertzog, S. N. Nair, W. A. Rogers, and J. Sharit. 2006. "Factors Predicting the Use of Technology: Findings from the Center for Research and Education on Aging and Technology Enhancement (CREATE)." *Psychology and Aging* 21 (2): 333–352. doi:10.1037/0882-7974.21.2.333.

Czaja, S. J., and J. Sharit. 2003. "Practically Relevant Research: Capturing Real World Tasks, Environments, and Outcomes." *The Gerontologist* 43 (suppl_1): 9–18. doi:10.1093/geront/43.suppl_1.9.

Davis, N. C., and D. Friedrich. 2010. "Age Stereotypes in Middle-Aged through Old-Old Adults." *The International Journal of Aging and Human Development* 70 (3): 199–212. doi:10.2190/AG.70.3.b.

de Graaf, M. M., and B. F. Malle. 2017. "How People Explain Action (and Autonomous Intelligent Systems Should Too)." In AAAI Fall Symposium on Artificial Intelligence for Human-Robot Interaction. Arlington, VA.

Endsley, M. R., and E. O. Kiris. 1995. "The out-of-the-Loop Performance Problem and Level of Control in Automation." *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37 (2): 381–394. doi:10.1518/001872095779064555.

Eyssel, F., and F. Hegel. 2012. "(S)He's Got the Look: Gender Stereotyping of Robots." *Journal of Applied Social Psychology* 42 (9): 2213–2230. doi:10.1111/j.1559-1816.2012.00937.x.

Eyssel, F., and D. Kuchenbrandt. 2012. "Social Categorization of Social Robots: Anthropomorphism as a Function of Robot Group Membership." *British Journal of Social Psychology* 51 (4): 724–731. doi:10.1111/j.2044-8309.2011.02082.x.

Fiske, S. T. 1993. "Controlling Other People: The Impact of Power on Stereotyping." *American Psychologist* 48 (6): 621–628. doi:10.1037/0003-066X.48.6.621.

Fiske, S. T. 1998. "Stereotyping, Prejudice, and Discrimination." In *The Handbook of Social Psychology*, edited by D. T. Gilbert and S. T. Fiske, 357–411. New York: McGraw-Hill.

Fiske, S. T. 2017. "Prejudices in Cultural Contexts: shared Stereotypes (Gender, Age) versus Variable Stereotypes (Race, Ethnicity, Religion)." *Perspectives on Psychological Science* 12 (5): 791–799. doi:10.1177/1745691617708204.

Fiske, S. T., and S. Taylor. 1991. *Social Cognition*. New York: McGraw-Hill.

Follett, K., and T. M. Hess. 2002. "Aging, Cognitive Complexity, and the Fundamental Attribution Error." *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences* 57 (4): P312–P323. doi:10.1093/geronb/57.4.P312.

Friemel, T. N. 2016. "The Digital Divide Has Grown Old: Determinants of a Digital Divide among Seniors." *New Media and Society* 18 (2): 313–331. doi:10.1177/1461444814538648.

Froese, T., and T. Ziemke. 2009. "Enactive Artificial Intelligence: Investigating the Systemic Organization of Life and Mind." *Artificial Intelligence* 173 (3–4): 466–500. doi:10.1016/j.artint.2008.12.001.

Gilbert, D. T., and P. S. Malone. 1995. "The Correspondence Bias." *Psychological Bulletin* 117 (1): 21–38. doi:10.1037/0033-2909.117.1.21.

Grossmann, I., and L. Ross. 2018. "Attributions." *The SAGE Encyclopedia of Lifespan Human Development* 1: 189–191.

Hale, N. M. 1998. “Effects of Age and Interpersonal Contact on Stereotyping of the Elderly.” *Current Psychology* 17 (1): 28–47. doi:10.1007/s12144-998-1019-2.

Hancock, P. A., D. R. Billings, K. E. Schaefer, J. C. Chen, E. J. de Visser, and R. Parasuraman. 2011. “A Meta-Analysis of Factors Affecting Trust in Human-Robot Interaction.” *Human Factors: The Journal of the Human Factors and Ergonomics Society* 53 (5): 517–527. doi:10.1177/0018720811417254.

Hayes, C. C., and C. A. Miller. 2010. *Human-Computer Etiquette: Cultural Expectations and the Design Implications They Place on Computers and Technology*. Boca Raton: CRC Press.

Hummert, M. L. 1994. “Physiognomic Cues and the Activation of Stereotypes of the Elderly in Interaction.” *The* International Journal of Aging and Human Development 39 (1): 5–20. doi:10.2190/6EF6-P8PF-YP6F-VPY4.

Hummert, M. L., T. A. Garstka, and J. L. Shaner. 1997. “Stereotyping of Older Adults: The Role of Target Facial Cues and Perceiver Characteristics.” *Psychology and Aging* 12 (1): 107–114. doi:10.1037/0882-7974.12.1.107.

Kaufmann, M. C., F. Krings, and S. Sczesny. 2016. “Looking Too Old? How an Older Age Appearance Reduces Chances of Being Hired.” *British Journal of Management* 27 (4): 727–739. doi:10.1111/1467-8551.12125.

Kelley, H. H. 1973. “The Processes of Causal Attribution.” *American Psychologist* 28 (2): 107–128. doi:10.1037/h0034225.

Kite, M. E., and B. T. Johnson. 1988. “Attitudes toward Older and Younger Adults: A Meta- Analysis.” *Psychology and Aging* 3 (3): 233–244. doi:10.1037/0882-7974.3.3.233.

Kite, M. E., G. D. Stockdale, B. J. Whitley, and B. T. Johnson. 2005. “Attitudes toward Younger and Older Adults: An Updated Meta-Analytic Review.” *Journal of Social Issues* 61 (2): 241–266. doi:10.1111/j.1540-4560.2005.00404.x.

Krull, D. S. 1993. “Does the Grist Change the Mill? the Effect of Perceiver’s Goals on the Process of Social Inference.” *Personality and Social Psychology Bulletin* 19 (3): 340–348. doi:10.1177/0146167293193011.

Krull, D. S., and D. J. Erickson. 1995. “Judging Situations: On the Effortful Process of Taking Dispositional Information into account.” *Social Cognition* 13 (4): 417–438. doi:10.1521/soco.1995.13.4.417.

Kuchenbrandt, D., M. Häring, J. Eichberg, F. Eyssel, and E. André. 2014. “Keep an Eye on the Task! How Gender Typicality of Tasks Influence Human–Robot Interactions.” *International Journal of Social Robotics* 6 (3): 417–427.

Lamont, R. A., H. J. Swift, and D. Abrams. 2015. “A Review and Meta-Analysis of Age-Based Stereotype Threat: Negative Stereotypes, Not Facts, Do the Damage.” *Psychology and Aging* 30 (1): 180–193. doi:10.1037/a0038586.

Lee, E. J. 2003. “Effects of “Gender” of the Computer on Informational Social Influence: The Moderating Role of Task Type.” *International Journal of Human-Computer Studies* 58 (4): 347–362. doi:10.1016/S1071-5819(03)00009-0.

Lee, J.D., and N. Moray. 1994. “Trust, Self-Confidence, and Operators’ Adaptation to Automation.” *International Journal of Human-Computer Studies* 40 (1): 153–184.

Levy, S. 2018. “Toward Reducing Ageism: PEACE (Positive Education about Aging and Contact Experiences) Model.” *The Gerontologist* 58 (2): 226–232. doi:10.1093/geront/gnw116.

Lineweaver, T. T., and C. Hertzog. 1998. “Adults’ Efficacy and Control Beliefs regarding Memory and Aging: Separating General from Personal Beliefs.” *Aging, Neuropsychology, and Cognition* 5 (4): 264–296. doi:10.1076/anec.5.4.264.771.

Muller-Johnson, K., M. P. Toglia, C. D. Sweeney, and S. J. Ceci. 2007. “The Perceived Credibility of Older Adults as Witnesses and Its Relation to Ageism.” *Behavioral Sciences and The Law* 25: 355–375. doi:10.1002/bsl.765.

Nass, C., J. Steuer, and E. R. Tauber. 1994. “Computers are Social Actors.” In Proceedings of the SIGCHI conference on Human factors in computing systems (pp. 72–78). Boston, MA: ACM. doi:10.1145/259963.260288.

Nass, C., and K. M. Lee. 2001. “Does Computer-Synthesized Speech Manifest Personality? Experimental Tests of Recognition, Similarity-Attraction, and Consistency-Attraction.” *Journal of Experimental Psychology: Applied* 7 (3): 171. doi:10.1037/1076-898X.7.3.171.

Nass, C., and Y. Moon. 2000. “Machines and Mindlessness: Social Responses to Computers.” *Journal of Social Issues* 56 (1): 81–103. doi:10.1111/0022-4537.00153.

Nass, C., Y. Moon, and N. Green. 1997. “Are Machines Gender Neutral? Gender Stereotypic Responses to Computers with Voices.” *Journal of Applied Social Psychology* 27 (10): 864–876. doi:10.1111/j.1559-1816.1997.tb00275.x.

Pak, R., N. Fink, M. Price, B. Bass, and L. Sturre. 2012. “Decision Support Aids with Anthropomorphic Characteristics Influence Trust and Performance in Younger and Older Adults.” *Ergonomics* 55 (9): 1059–1072. doi:10.1080/00140139.2012.691554.

Pak, R., A. McLaughlin, and B. Bass. 2014. “A Multi-Level Analysis of the Effects of Age and Gender Stereotypes on Trust in Anthropomorphic Technology by Younger and Older Adults.” *Ergonomics* 57 (9): 1277–1289. doi:10.1080/00140139.2014.928750.

Pak, R., E. Rovira, A. C. McLaughlin, and N. Baldwin. 2017. “Does the Domain of Technology Impact User Trust? Investigating Trust in Automation across Different Consumer-Oriented Domains in Young Adults, Military, and Older Adults.” *Theoretical Issues in Ergonomics Science* 18 (3): 199–220. doi:10.1080/1463922X.2016.1175523.

Palumbo, R., R. J. Adams, U. Hess, R. E. Kleck, and L. Zebrowitz. 2017. “Age and Gender Differences in Facial Attractiveness, but Not Emotion Resemblance, Contribute to Age and Gender Stereotypes.” *Frontiers in Psychology* 8: 1–18. doi:10.3389/fpsyg.2017.01704.

Parasuraman, R., and C. Miller. 2004. “Trust and Etiquette in High-Criticality Automated Systems”. *Communications of the ACM* 47 (4): 51–55. doi:10.1145/975817.975844.

Reeves, B., and C. I. Nass. 1996. *The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places*. Cambridge, England: Cambridge university press.

Richardson, J. 2011. “Eta Squared and Partial Eta Squared as Measures of Effect Size in Educational Research.” *Educational Research Review* 6 (2): 135–147. doi:10.1016/j.edurev.2010.12.001.

Rossi, P. H., and A. B. Anderson. 1982. “The Factorial Survey Approach: An Introduction.” *Measuring Social Judgments.*

The Factorial Survey Approach, edited by P. H. Rossi & S. L. Nock. 15–67. Beverly Hills, CA: SAGE Publications.

Ruffman, T., J. Murray, J. Halberstadt, and T. Vater. 2012. "Age-Related Differences in Deception." *Psychology and Aging* 27 (3): 543–549. doi:10.1037/a0023380.

Schaefer, K. E., T. L. Sanders, R. E. Yordon, D. R. Billings, and P. A. Hancock. 2012. "Classification of Robot Form: Factors Predicting Perceived Trustworthiness." *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56 (1): 1548–1552. doi:10.1177/1071181312561308.

Schaich, A., S. Obermeyer, T. Kolling, and M. Knopf. 2016. "An Own-Age Bias in Recognizing Faces with Horizontal Information." *Frontiers in Aging Neuroscience* 8: 264–274. doi:10.3389/fnagi.2016.00264.

Shah, R., and J. Ogden. 2006. "'What's in a Face?'The Role of Doctor Ethnicity, Age and Gender in the Formation of Patients' Judgements: An Experimental Study." *Patient Education and Counseling*, 60 (2): 136–141.

Sharkey, A., and N. Sharkey. 2012. "Granny and the Robots: ethical Issues in Robot Care for the Elderly." *Ethics and Information Technology* 14 (1): 27–40. doi:10.1007/s10676-010-9234-6.

Sherman, J. W. 2001. "The Dynamic Relationship between Stereotype Efficiency and Mental Representation." In G. Moskowitz (Ed.), *Cognitive Social Psychology: The Princeton Symposium on the Legacy and Future of Social Cognition*, 177–190. Mahwah NJ: Erlbaum.

Singh, I. L., R. Molloy, and R. Parasuraman. 1993. "Automation-Induced "Complacency": Development of a Complacency-Potential Scale." *The International Journal of Aviation Psychology* 3 (2): 111–122. doi:10.1207/s15327108ijap0302_2.

Sung, J., E. Grinter, H. I. Christensen, and L. Guo. 2008. "Housewives or Technophiles? Understanding Domestic Robot Owners." Proceedings of the 3rd ACM/IEEE Conference on Human-Robot Interaction, 2008. HRI'08, Amsterdam, The Netherlands, March 12–15, 2008. ACM, New York, 129–136.

Tay, B., Y. Jung, and T. Park. 2014. "When Stereotypes Meet Robots: The Double-Edge Sword of Robot Gender and Personality in Human–Robot Interaction." *Computers in Human Behavior* 38: 75–84. doi:10.1016/j.chb.2014.05.014.

Tomlinson, E. C., and R. C. Mayer. 2009. "The Role of Causal Attribution Dimensions in Trust Repair." *Academy of Management Review* 34 (1): 85–104. doi:10.5465/amr.2009.35713291.

Trope, Y. 1986. "Identification and Inferential Processes in Dispositional Attribution." *Psychological Review* 93 (3): 239–257. doi:10.1037/0033-295X.93.3.239.

Von Hippel, W., L. A. Silver, and M. E. Lynch. 2000. "Stereotyping against Your Will: The Role of Inhibitory Ability in Stereotyping and Prejudice among the Elderly." *Personality and Social Psychology Bulletin* 26 (5): 523–532. doi:10.1177/0146167200267001.

Wiese, H., J. Komes, and S. R. Schweinberger. 2013. "Ageing Faces in Ageing Minds: A Review on the Own-Age Bias in Face Recognition." *Visual Cognition* 21 (9–10): 1337–1363. doi:10.1080/13506285.2013.823139.

Wortham, R. H., A. Theodorou, and J. J. Bryson. 2017. "Improving Robot Transparency: real-Time Visualisation of Robot AI Substantially Improves Understanding in Naive Observers." In *Robot and Human Interactive Communication (RO-MAN), 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1424–1431. Lisbon, Portugal: IEEE.

Xu, Q., J. Ng, O. Tan, Z. Huang, B. Tay, and T. Park. 2015. "Methodological Issues in Scenario-Based Evaluation of Human–Robot Interaction." *International Journal of Social Robotics* 7 (2): 279–291. doi:10.1007/s12369-014-0248-9.

Zickuhr, K. 2011. Generations and their gadgets. Pew Internet and American Life Projec. Retrieved from https://www.pewresearch.org/internet/2011/02/03/generations-and-their-gadgets/.

Ziemke, T., S. Thill, and D. Vernon. 2015. "Embodiment is a Double-Edged Sword in Human-Robot Interaction: Ascribed vs. Intrinsic Intentionality." Proceedings of the Workshop on Cognition: A Bridge between Robotics and Interaction, ACM/IEEE Human Robot Interaction Conference (HRI), Portland, OR, 9–10.