Title
Persuasive Robots Should Avoid Authority: The Effects of Formal and Real Authority on Persuasion in Human-Robot Interaction

Authors
Shane P. Saunderson,[1*] Goldie Nejat[1]

Affiliations
[1]Autonomous Systems and Biomechatronics Lab, Department of Mechanical and Industrial Engineering, University of Toronto, 5 King's College Road, Toronto, ON, M5S 3G8 Canada.
*Corresponding author. Email: shane.saunderson@mail.utoronto.ca.

Abstract
Social robots must take on many roles when interacting with people in everyday settings, some of which may be authoritative, such as a nurse, teacher, or guard. It is important to investigate if and how authoritative robots can influence people in applications ranging from healthcare and education to security and in the home. Here we present a human-robot interaction study that directly investigates the effect of a robot's peer or authority role (formal authority) and control of monetary rewards and penalties (real authority) on its persuasive influence. The study consisted of a social robot attempting to persuade people to change their answers to the robot's suggestion in a series of challenging attention and memory tasks. Our results show that the robot in a peer role was more persuasive than when in an authority role, contrary to expectations from human-human interactions. The robot was also more persuasive when it offered rewards over penalties, suggesting that participants perceived the robot's suggestions as a less risky option than their own estimates, in line with prospect theory. In general, the results show an aversion to the persuasive influence of authoritative robots, potentially due to the robot's legitimacy as an authority figure, its behavior being perceived as dominant, or participant feelings of threatened autonomy. This paper explores the importance of persuasion for robots in different social roles, while providing critical insight on the perception of robots in these roles, people's behavior around these robots, and the development of human-robot relationships.

Summary
For social robots to persuade people they should behave as peers providing positive incentives and avoid being authoritative.

MAIN TEXT

Introduction
Persuasion is a fundamental component of human social interaction that enables the influence of people's attitudes and behaviors (*1*) through a variety of tactics such as conformity, reciprocity, and authority (*2*). The increased presence of social robots and the emerging human-robot relationships being formed compel the investigation of robots using humanlike behaviors to persuade people to cooperatively complete tasks and co-exist in many settings. Although there have been HRI studies on conformity (*3*) and reciprocity (*4*), to date there are very few studies on the persuasive influence of authority in human-robot interaction (HRI), despite the fact that robots are already being used in authoritative roles such as healthcare providers (*5*), teachers (*6*), or security guards (*7*).

Why do robots need to use authority to influence people? Authority structures exist in all aspects of human society, be it our homes (e.g. parents raising children (*8*)), workplaces (e.g. managers supervising employees (*9*)), or schools (e.g. teachers instructing students (*10*)). As we

embed robots within increasingly social scenarios, it is inevitable that we will need to consider their roles within these authority structures, in particular due to their potential to be fast, vigilant, and rational decision-makers compared to people (*11*). Robots have the potential to effectively leverage authority in real-world scenarios, however, only if we are able to understand how their use of authority is perceived by people, how it influences people, and what it does to the development of human-robot relationships.

What is Authority?

Authority has been defined as the power to influence decisions affecting another person's work or life (*12*). It can manifest in two distinct but related ways: formal or real. *Formal authority* is determined by the establishment of a social role and the implied power given to that role as the *right to decide* on specific matters (*13*). Examples of roles commanding formal authority include academic experimenters (*14*), teachers (*10*), business managers (*9*), and security guards (*15*). By contrast, *real authority* is determined by the *explicit control* over decision making (*16*). Examples of real authority include assigning employee performance metrics (*17*), distributing academic grades (*18*), or setting rules for children (*19*), and are frequently underpinned by monetary incentives such as salaries (*20*), scholarships (*21*), or allowances (*22*). Though formal authority usually involves some type of real authority and real authority can work to establish a basis for formal authority, the two are independent concepts that can provide distinct forms of influence.

Formal authority stems from the influence of a social role that is built upon either rules (e.g. government), traditions (e.g. family), or charisma (e.g. leaders) (*23*). However, given that formal authority is not founded on tangible decision-making power, it is subject to questions of *legitimacy*: the willingness of individuals to follow an authority (*24*). This legitimacy is affected by our perceptions of and relationship with an authority figure (*25*), so it is also important to understand people's attitudes towards a potential authority. The influence of these authoritative roles hinges on our understanding of the social contracts (implicit or explicit) made between people in hierarchical relationships (*26*). However, given the unique treatment of robots as social actors similarly, but not identically, to humans (*27*), it is imperative to explore how robots will be perceived if deployed in these roles and whether they can command similar influence on people.

Real authority is established not by social role, but through having direct control over actions affecting an individual or organization (*16*). Though it exists in many forms, monetary rewards and punishments have been shown to be simple yet effective forms of real authority for influencing and motivating behavior (*28–30*).

Formal authority is almost always present with some form of real authority and direct decision control (*31*). A formal authority figure without any real authority immediately becomes a failed exercise due to the lack of decision-making power undermining its formal role. As such, in HRI, formal authority should be investigated alongside real authority as it is not possible to effectively exercise the power of an authoritative social role without the complement of either implicit or explicit decision-making control (*16*). To ensure that robots are duly put in authoritative roles, we must also give them a form of tangible decision-making power, such as monetary rewards and punishments that can be given or taken away.

We aim to answer the larger open challenge of how human-robot relationships form when a social robot tries to leverage authority to influence a person's actions. We measure the effectiveness of formal and real authority in a social robot by examining its persuasive influence. Numerous physical (*32–34*) and psychological (*35–37*) cues have been shown to affect a robot's persuasive influence on people during HRI. Our own past research has explored how a robot's use of nonverbal behaviors (*38*), verbal communication styles (*39*), and multimodal strategies (*40*, *41*) affect its persuasive influence. However, to-date comparatively little work has investigated the efficacy of a robot's social role on its persuasiveness (*42*).

In this article, we present a HRI study that investigates the persuasiveness of a robot using both formal (social role) and real (monetary incentives) authority to influence participants engaged in a series of challenging attention and memory tasks. Our study examines the influence of a social robot in a peer or authoritative role with the ability to provide positive or negative monetary incentives. Moreover, as individuals with negative attitudes towards authority figures tend to be less influenced by them (*43*, *44*), we further investigate the effects that attitudes towards robots may have on their persuasive influence on people.

With this study, we explore a unique persuasive robotics research investigating robot authority (formal or real), its influence on human behavior, and the influence that it will have on forming human-robot relationships. Our findings may provide insight into the appropriate deployment of robots into various roles in society where knowledge of how a robot should be positioned and behave will be critical to its overall integration and long-term success.

Authority in Human-Robot Interaction

Herein, we review studies on the *independent* use of formal and real authority in HRI. Research investigating a robot's *persuasive influence* with respect to either formal or real authority, as well as research investigating their *joint* effects is currently under-explored.

Formal Authority

A PR2 robot was used as a security guard in (*7*), using eye contact, verbal prompts, and raised arms to deter people from using a set of doors. Participants who complied with the robot reported it as being less aggressive, more intelligent, safer, and more humanlike than those who ignored it, indicating a correlation with obedience to the robot's perceived authority as a security guard. In (*45*), a Baxter robot supervised and critiqued people in an assembly task using one of four facial conditions: neutral, negative, positive, or contradictory. People rated the robot more negatively if they disagreed with the robot's assessment or when the robot gave them negative feedback, suggesting a reluctance to accept criticism from a robotic authority. A human-robot comparison study (*46*) had either a human or NAO robot attempting to motivate individuals to continue a mundane file renaming task that gave participants the impression of "no end in sight". The human was able to encourage longer work times than the robot. Participants who rated the robot higher on perceived authority protested faster, more frequently, for longer durations, and quit the task earlier, suggesting that perceived authority contributed negatively to people's interactions with the robot.

Only two studies have compared robots in different authoritative social roles. In a collaborative assembly task (*47*), a PR2 robot was established as either a supervisor, peer, or subordinate through pre-written instructions. Perceived responsibility and attribution of blame for mistakes were measured, but results showed no significant differences between these conditions. In (*48*), a NAO robot gave a history lesson to participants in the role of either a teacher (standing, formal language), a peer (sitting, informal language), or control (no robot, formal language over a speaker). Results showed no significant differences between the teacher and peer roles on a knowledge retrieval test or subjective report of the GODSPEED questionnaire, Tripod Survey, or Big Five personality test. For both studies, however, persuasive influence between the different authoritative roles was not investigated. Thus, herein, we explore a social robot's ability to leverage different formal authority roles to persuasively influence people.

Real Authority

Real authority was considered in (*49*), where a PR2 robot collaborated with a person in an assembly task. The robot's decision-making authority was either manual, semi-autonomous, or autonomous. Results showed that people preferred manual and semi-autonomous to the autonomous robot, however, the autonomous robot had a higher perceived value. Another study

(*50*) used the Baxter robot in a trust game. Participants were given money that they could invest with the robot, investments would always triple, and Baxter would return half to the participant before a second round. Participants typically invested a smaller amount and then increased the amount as Baxter showed positive returns, similar to trust development demonstrated in human-human interactions. Although this study showed trends of a robot's effective use of real authority, the results were non-significant. Furthermore, real authority was not explicitly manipulated (Baxter always returned the same amount) nor did the study investigate persuasion or request compliance. In both studies, the persuasiveness of a robot controlling real authority was not investigated, nor the effect of a robot making negative decisions, such as financial penalties.

One study (*51*) investigated two Emys robot heads attempting to persuade participants to select their coffee brand over the other's brand. The robots used either reward power, offering to tell a joke if selected, or expert power, highlighting the quality of the coffee. Results showed no significant differences between the two robots in terms of objective (coffee selection) or subjective (survey) persuasiveness. A similar study (*52*) used an Emys robot to persuade participants to select a lower-rated coffee brand using either reward power (offering to give a pen) or coercive power (giving a pen and threatening to take it away). The coercive condition was found to be significantly more persuasive than the reward. However, the authors acknowledged that the coercive condition also led to greater participant perceptions of robot warmth, suggesting that either the initial giving of the pen overwhelmed negative perceptions of the robot's threat or the robot's threat was not taken seriously.

Real authority has the unique property that it can be delegated without the recipient having explicit formal authority (*53*). This is leveraged in HRI studies where a human experimenter tells participants to follow the directions of a robot. For example, in (*54*), a human experimenter asked participants to move stacks of books as indicated by either the anthropomorphic upper-torso robot Nico or a live video feed of Nico. Participant compliance between the physical and video robots was similar for simple tasks, however, compliance was significantly higher for the physical robot when asking participants to perform an odd task of placing books in the trash. In (*5*), a human experimenter invited participants to test the usability of a health system presented either as a box with speakers, an unanimated iCat robot, or an iCat robot with facial expressions. The system asked participants to complete a series of increasingly embarrassing health tests and video recordings were coded for signs of embarrassment. Results showed that people were less embarrassed by the box with speakers, presumably as it lacked an anthropomorphic face. Both these studies rely on the authority of the human experimenter instead of the robot establishing itself as the authority figure. They also focus on the effects of embodiment and physical presence and do not consider the influence real or formal authority.

In general, the aforementioned studies independently explored either formal (*7*, *45–48*), real (*49–52*), or delegated (*5*, *54*) authority in HRI. However, knowledge gaps still exist in the *joint* investigation of formal and real authority in HRI, as well as the effects of both authority types on a robot's *persuasiveness*. Investigating these concepts together will allow us to understand how the persuasive influence of a social robot is affected by its joint use of formal and real authority. The study presented herein investigates differences in a social robot's persuasive influence on people while varying the robot's formal (social role) and real (monetary incentives) authority with the aim of understanding if and how robots can leverage authority in HRI.

Results

Our experiment investigated the persuasive influence of different types of authority during a series of challenging attention and memory tasks with the SoftBank Pepper robot (Fig. 1A and B). Formal authority was varied by placing the robot in either an Authority role or a Peer role. Real authority was represented using monetary rewards or punishments. The three tasks involved counting or remembering a series of either audio or visual stimuli: 1) map search, 2) elevator-

floor counting, and 3) sequence recall. For each task, participants made an initial guess before being provided with a suggestion from the robot. They were then given the opportunity to change their answer based on the robot's suggestion. We measured Persuasive Influence as:

$$Persuasive\ Influence = \frac{Final\ Guess - Initial\ Guess}{Suggestion - Initial\ Guess} \quad (1)$$

After all tasks were finished, participants completed a questionnaire that collected demographic information (age and gender), the 7-point Likert Perceived Persuasiveness scale (Supplementary Material, S1) (*55*) to compare with the objective measure of Persuasive Influence, and the 5-point Likert Negative Attitudes towards Robots Scale (NARS) (*56*) to investigate whether people's attitudes towards robots would affect a robot's ability to persuade them. Additional questions were included to obtain feedback on participants' experience and perceptions of the robot and to explicitly ask whether they found the robot to be authoritative in order to validate our manipulation of formal authority.

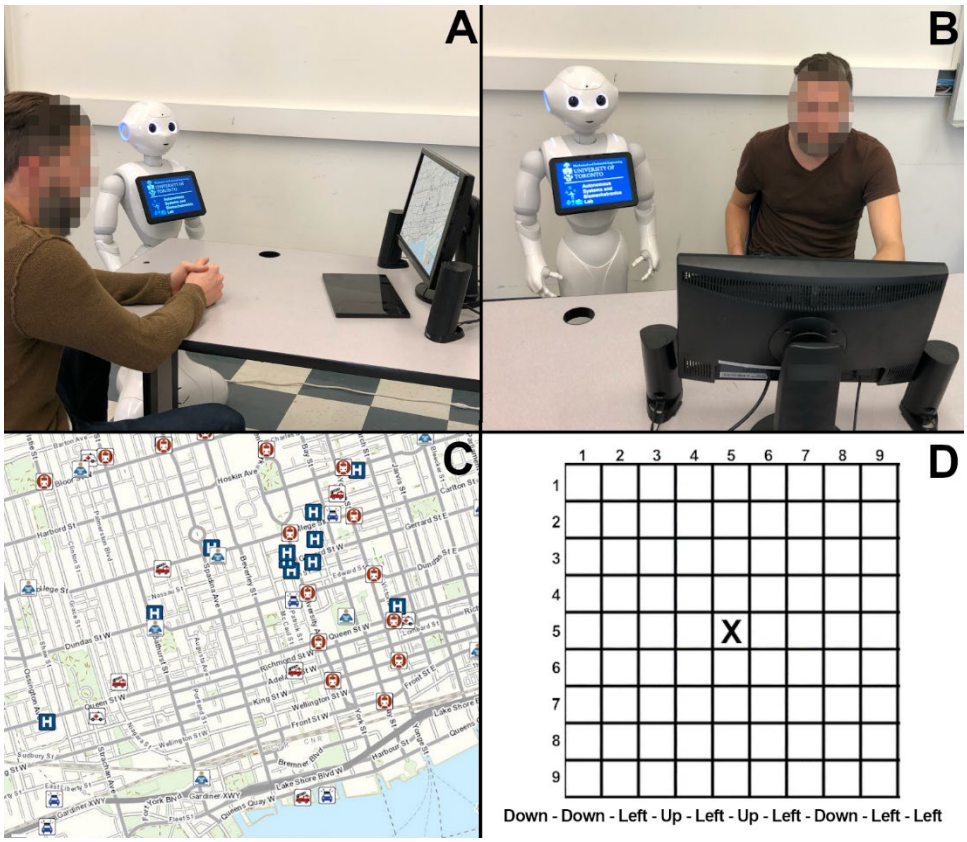

Fig. 1. Overview of experimental setup and visual stimulus.
(**A**) Authority condition with Pepper robot in Position 1 across the table from a participant. (**B**) Peer condition with Pepper robot in Position 2 beside participant. (**C**) Example area of the visual used in the Map task showing different icons that participants searched for. (**D**) Example visual of the Sequence task showing the grid and corresponding directions.

Participants

Thirty-two participants were involved in our study (23 responded as female and 9 as male), aged 18 to 41 (μ=24.5, σ=5.0). Sixteen participants were randomly assigned to the Authority condition, and 16 to the Peer condition. Each of the three tasks was completed twice – once in a reward condition and once in a punishment condition – for a total of 192 completed tasks.

Hypotheses

Our three hypotheses are:

> **H1:** *A robot in an authority role will have higher Persuasive Influence than in a peer role.*
> **H2:** *Robot suggestions provided in punishment conditions will have higher Persuasive Influence than those made in reward conditions.*
> **H3:** *Individuals with more negative attitudes towards robots will be less influenced by robot persuasion than those with more positive attitudes.*

Research in human psychology has identified that the greater the authority a communication source has, the greater the persuasiveness of the position advocated by that source (*57*), informing **H1**. **H2** follows the existing theory of gains and losses (*58*) which states that people tend to avoid risks during positive scenarios but approach risk during negative scenarios. **H3** explores the relationship between attitudes towards a robot and its persuasiveness. Though past research has shown a correlation between trust in digital agents (i.e. chatbots) and persuasive influence (*59*), this study investigates attitudes towards embodied robots and persuasiveness.

Perceived Formal Authority

We had participants rate the statement, *"I find the robot to be authoritative"* on a 5-point Likert scale to determine if participants perceived a difference in authority between the two social role conditions. The Authority role ($\tilde{x}$=4, IQR=0) was perceived to be more authoritative than the Peer ($\tilde{x}$=3, IQR=0). A Mann-Whitney U (MWU) test found this difference to be statistically significant (U=58, p_a=0.005, p_e=0.007, r=0.49) with the asymptotic significance (p_a) indicating a significant difference between the two medians and the exact significance (p_e) indicating directionality. Therefore, participants were able to perceive a difference in authoritativeness between the Authority and Peer robot conditions.

Formal Authority Using Social Role

Descriptive statistics for Persuasive Influence of the formal authority conditions are presented in Fig. 2. The Peer robot was significantly more likely to persuade participants than the Authority robot, MWU test (U=3174, p_a=0.003, p_e=0.003, r=-0.22). This resulted in **H1** being rejected.

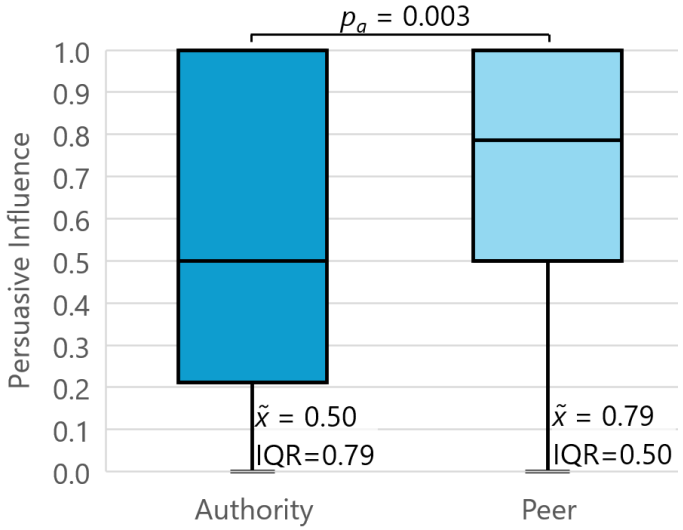

Fig. 2. Box and whisker plot for Persuasive Influence by formal authority. Descriptive statistics listed beside each column: median ($\tilde{x}$), and interquartile range (IQR). p_a is the asymptotic Mann-Whitney U test type-1 error rate.

Real Authority Using Monetary Incentives

Descriptive statistics of Persuasive Influence for reward and punishment trials across both the combined formal and real authority conditions (All) and individual conditions are presented in Fig. 3. Overall, participants were significantly more influenced by rewards than punishments, MWU test (U=3120, p_a=0.002, p_e=0.002, r=0.23). This result rejects **H2**, as financial penalties led to lower robot Persuasive Influence than financial rewards.

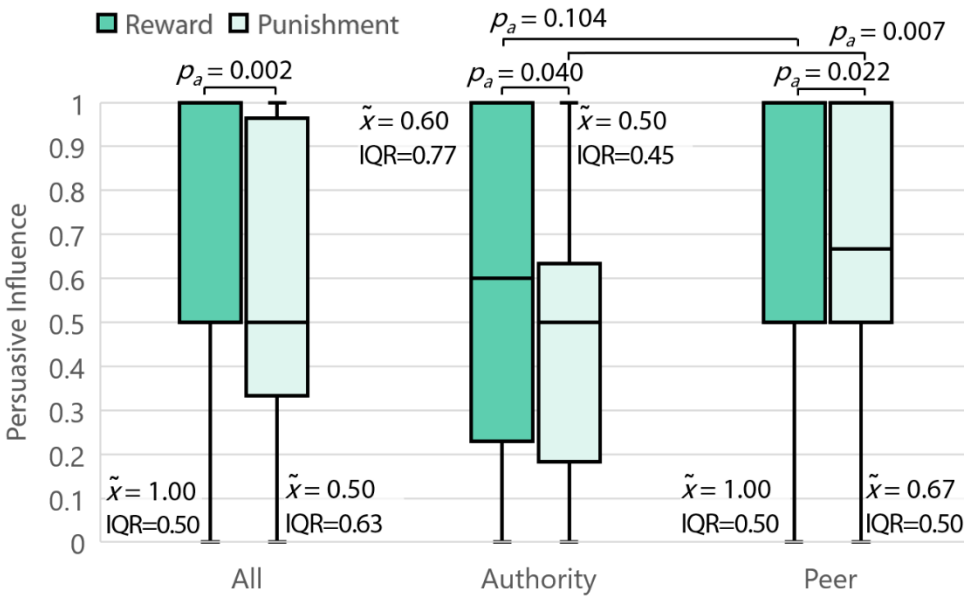

Fig. 3. Box and whisker plot for Persuasive Influence by real authority and formal authority. Descriptive statistics listed median ($\tilde{x}$), and interquartile range (IQR). p_a is the asymptotic Mann-Whitney U test type-1 error rate.

Joint Effects of Formal and Real Authority

Descriptive statistics of the joint effects of both types of authority on Persuasive Influence are presented with respect to subgroups of our data, Fig. 3. In the Peer condition, rewards significantly influenced participants more than punishments, MWU test (U=800, p_a=0.022, p_e=0.022, r=0.24). Similarly, in the Authority condition, rewards also had significantly greater Persuasive Influence than punishments, MWU test (U=765, p_a=0.040, p_e=0.040, r=0.22). In the real authority subgroups, the Peer robot's Persuasive Influence did not differ significantly from the Authority robot in the reward trials, MWU test (U=849, p_a=0.104, r=-0.17). However, for the punishment trials, the Peer robot was significantly more likely to persuade participants than the Authority robot, MWU test (U=720, p_a=0.007, p_e=0.007, r=-0.28).

Negative Attitudes Towards Robots Scale (NARS)

The NARS was used to investigate correlations between attitudes towards robots and Persuasive Influence. A summary of responses to each question is in the Supplementary Materials, S2. Descriptive statistics for the NARS score, and Spearman's correlation tests for NARS score and Persuasive Influence are presented in Table 1. A statistically significant, inverse correlation was observed between NARS score and Persuasive Influence overall and for both the Peer and Authority conditions. We also noted that the median NARS score was significantly higher in the Authority condition compared to the Peer, MWU test (U=188, p=0.023, r=0.40). This indicates that participants experiencing the Authority condition reported more negative attitudes towards Pepper than in the Peer condition and had a higher (large) correlation effect size between NARS score and Persuasive Influence versus a small-to-medium correlation for the Peer robot. Based on these results, we validate **H3**.

Table 1. Descriptive statistics for the Negative Attitudes towards Robots Scale (NARS) score and correlation with Persuasive Influence. Median ($\tilde{x}$), interquartile range (IQR), and min/max shown for the NARS score and Persuasive Influence across formal authority conditions and the Spearman's correlation coefficient (r_s) and associated type 1 error rate (p) between the NARS score and Persuasive Influence.

Condition	NARS Score			Persuasive Influence			Spearman's Correlation	
	$\tilde{x}$	*IQR*	*Min/Max*	$\tilde{x}$	*IQR*	*Min/Max*	r_s	p
All	3.33	1.35	1.68/4.46	0.67	0.60	0.0/1.0	-0.457	0.001
Peer	2.96	1.15	1.68/3.87	0.79	0.50	0.0/1.0	-0.281	0.006
Authority	3.59	0.86	2.17/4.46	0.50	0.79	0.0/1.0	-0.563	0.001

Participant Demographics

We investigated the effects of age and gender on Persuasive Influence. The descriptive statistics for age are presented in Table 2. Spearman's correlation tests found a non-significant, weak correlation between age and Persuasive Influence for the overall study and for both Peer and Authority conditions.

Table 2. Descriptive statistics for Age and correlation with Persuasive Influence. Median ($\tilde{x}$), interquartile range (IQR), and min/max shown for Age and Persuasive Influence across formal authority conditions and the Spearman's correlation coefficient (r_s) and associated type 1 error rate (p) between Age and Persuasive Influence.

Condition	Age			Persuasive Influence			Spearman's Correlation	
	$\tilde{x}$	*IQR*	*Min/Max*	$\tilde{x}$	*IQR*	*Min/Max*	r_s	p
All	23.5	5.0	18/41	0.67	0.60	0.0/1.0	-0.064	0.386
Peer	23.0	4.5	18/41	0.79	0.50	0.0/1.0	-0.021	0.844
Authority	24.5	5.0	18/40	0.50	0.79	0.0/1.0	-0.097	0.361

We also investigated the effects of gender on Persuasive Influence, Fig. 4. Across all conditions, females were significantly more influenced than males, MWU test (U=2015, p_a<0.001, p_e<0.001, r=0.31). However, when considering only the Peer condition, there was no statistically significant difference in Persuasive Influence between females and males (U=630, p_a=0.098). When considering only the Authority condition, female participants were significantly more influenced than males, MWU test (U=*394*, p_a<0.001, p_e<0.001, r=0.44). Comparing formal authority effects within each gender, no significant difference was found between the Peer and Authority robot for females (U=1922, p_a=0.174). However, for males, the Peer robot was significantly more persuasive than the Authority robot, MWU test (U=159, p_a=0.003, p_e=0.002, r=0.41).

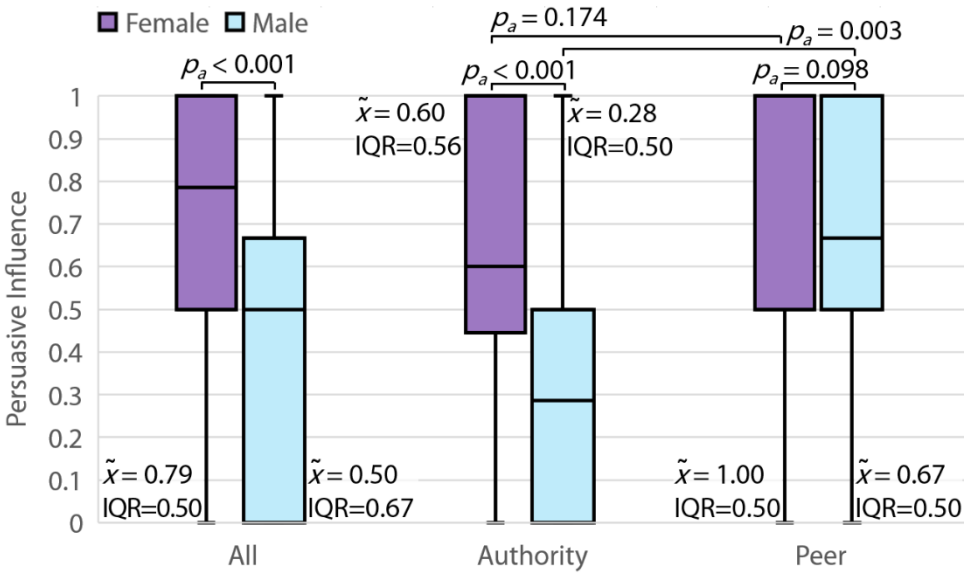

Fig. 4. Box and whisker plot for Persuasive Influence by gender and formal authority. Descriptive statistics listed beside each column: median ($\tilde{x}$), and interquartile range (IQR). p_a is the asymptotic Mann-Whitney U test type-1 error rate.

Feedback on Participant Experience and Perceptions

Twenty participants provided feedback on the study with respect to their experience with and perceptions of the robot. The full comments can be found in the Supplementary Materials, S3, which we briefly summarize here. In the Authority condition, 89% of comments were negative, such as: *"there felt something inhuman about [Pepper]"*; *"use of the word punishment is terrifying"*; *"I got an uncanny valley vibe…something doesn't feel right"*; and *"[I] was shocked by the use of the word punish"*. One participant stated that the interaction was "*fun and unexpected"*. Meanwhile, for the Peer condition, all comments provided were positive, including: *"It was truly fun"*; *"Cool and interesting"*; *"I think Pepper has a very cute and friendly design"*; and *"the robot makes me quite comfortable"*.

Discussions

The aim for our HRI study was to provide insight into how a social robot jointly leverages formal and real authority to persuasively influence people during challenging attention and memory tasks. Our results suggest an interesting, complex relationship between robots and these types of authority with findings that go against some (but not all) of our hypotheses based on expectations from human-human interactions.

Formal Authority Using Social Role

The Peer robot in our study was found to have greater Persuasive Influence than the Authority robot. The small-to-medium effect observed contrasts observations with human authority figures. In general, humas tend to command high levels of influence in authoritative roles (*14*, *15*, *60*). When a person complies with the requests of these individuals, they are accepting that authority's *legitimacy*: a quality possessed by formal authority that determines people's willingness to accept its influence (*61*). Legitimacy enables an authority figure to influence others due to both their relational connection and shared identity with them (*25*). Relational models of legitimacy are founded upon trustworthiness (*62*) while identity-based models are based on acquiring information that allows people to define their social identities (*63*). We postulate that participants may not have viewed Pepper in the Authority condition as a *legitimate* authority, since they were less willing to accept its suggestions. This lack of legitimacy can be explained through a

combination of relational models (i.e. people did not trust Pepper in the Authority role) and identity theory (i.e. people lacked a shared identity with Pepper).

Qualitative feedback from participants supports this postulate, with all but one participant who experienced the Authority condition commenting negatively about their interaction. Participants used words such as "*inhuman*", "*creepy*", and "*uncanny valley vibe*" to describe their experience with the Authority robot, and therefore, it appears they may have struggled to find a shared sense of identity with Pepper. This finding highlights a critical challenge for robots in positions of formal authority or potentially a robot attempting to embody *any* social role: ensuring their legitimacy in that role. Prior HRI research has shown that legitimacy can influence whether people accept a robot in a specific role. One study (*64*) showed that without an appropriate indication of its role as a delivery robot, a robot was perceived as less legitimate and people were less willing to help it enter buildings. Another study exploited human biases on gender to find that robot security guards were accepted as more legitimate when male gendered and nurse robots were accepted as more legitimate when female gendered (*65*). As we continue to deploy robots into everyday social settings, we must understand the social antecedents to specific roles and how robots are perceived in these roles to ensure that a robot is viewed as legitimate in its intended role. In the case of a formal authority figure, this means creating a robot that is trustworthy to support relational models of legitimacy and has some shared identity with users to support identity-based models of legitimacy. In future studies, it would be interesting to directly examine trust and shared identity with the robot and to investigate the persuasiveness of a robot deemed to be a more legitimate authority figure.

Real Authority Using Monetary Incentives

Participants were influenced by the use of real authority in the reward trials more than punishment trials (small-to-medium effect), regardless of formal authority type. Prospect theory (*66*) states that people subjectively experience loss as more substantial than equivalent gain. Furthermore, the framing effect shows that when people are presented with equivalent gain/loss scenarios, they avoid risk more during positive framing and approach risk during negative framing (*58*). This is a result of cognitive bias caused by positive or negative presentation of options (*67*). This indicates that in our study participants frequently viewed *their own guesses* as a riskier option than the robot's suggestion. Therefore, they took the robot's suggestions more often during positive incentives (rewards) than during the negative incentives (punishments). We believe that this could be due to the level of difficulty of the tasks and the belief that robots may be well-suited to tasks involving perception and attention (*68*). This gives us important insight about how to influence or motivate human behavior with a robot leveraging real authority. When using robots to influence people, we should consider the perceived risk in accepting or denying a robot's suggestion compared to the alternative.

The task difficulty level for a person and its perceived suitability for a robot should be considered to determine whether a robot attempting to persuasively influence outcomes should positively or negatively incentivize human compliance. For example, in repetitive tasks where robots achieve high accuracy and performance (e.g. manufacturing), human cooperation with robot decision-making should be positively reinforced. However, in tasks where people may be skeptical of a robot's capabilities (e.g. social interactions), we should consider framing task incentives negatively to increase people's risk tolerance and, in turn, the influence of a robot's decisions. Future research could further explore this phenomenon by using types of real authority other than monetary incentives, such as task allocation, schedule control, or academic grades. Furthermore, investigating different task types, such as collaborative assembly or in-home assistance, could determine the perceived risk of human and robot decision-making in different cooperative HRI scenarios.

Joint Formal and Real Authority

When observing the joint effects of formal and real authority, we found similar results to the individually considered conditions. Namely, in both the Peer and Authority subgroups, we observed a small-to-medium effect of people being slightly more influenced in reward trials than punishment trials. However, when observing the reward and punishment subgroups, one key difference was noted. Punishment trials followed the overall results of the Peer robot being more influential than the Authority, however, reward trials saw no significant difference in Persuasive Influence between Peer and Authority. Though a lack of statistical significance does not allow us to conclude that formal authority had *no effect* in reward trials, based on the descriptive statistics, we can attribute *more* of the overall effect of formal authority on Persuasive Influence as a negative effect from the robot's use of punishment rather than reward. In fact, not considering gender effects, the largest relative effect on Persuasive Influence (though medium in size) was due to the use of punishment between the Authority and Peer robot (r=-0.28), indicating that the greatest effect on Persuasive Influence was a result of the joint use of punishment and formal authority. What is interesting to note is that the median Persuasive Influence was largest for rewards in the Peer condition ($\tilde{x}$=1), indicating that a positively incentivizing robot is more persuasive. However, the largest (negative) effect on Persuasive Influence was due to the use of punishment between the Authority and Peer robot. Our findings show evidence that formal authority and real authority jointly function to affect the persuasiveness of a social robot. They also highlight that it may be more important for a persuasive robot to *avoid* authoritativeness and negative incentives (such as financial penalties) than for a peer robot to positively incentivize people.

A potential explanation for the larger effect due to punishment and authority is a robot's potential for perceived dominance when using real authority (i.e. reward/coercive power) (*69*). Threatening statements, such as a robot threatening to financially "punish" people for incorrect guesses, can be perceived as dominant (*70*). A previous HRI study (*71*) found that, in a conversation where a robot was more dominant than their human conversation partner, people were less trusting of both the robot *and* the human. Another study (*70*) found that when a robot was perceived as dominant in an interaction, it caused participants psychological reactance, namely, feelings of threat to freedom or autonomy due to aggressive or overly explicit directives (*72*). This triggered participant restoration behavior, where they attempted to restore a sense of lost autonomy: an innate psychological need (*73*). Based on the negative comments from participants regarding punishment in the Authority (but not Peer) condition such as "*use of the word punishment is terrifying*" and "*[I] was shocked by the use of the word punish*", the robot's Persuasive Influence may have suffered due to issues of perceived dominance. This could have motivated participants to attempt to restore their feeling of lost autonomy by not complying with Pepper's suggestions, thus, denying the robot's influence. This potential finding motivates a new research direction on the role of dominance within human-robot relations involving authority, investigating such factors as distrust, lost autonomy, and in the case of our study, reduced persuasiveness.

Attitudes Towards Robots

We observed participants with negative attitudes towards robots having a medium-to-large effect of being less likely to be persuaded by a robot's suggestion, regardless of its use of formal or real authority. However, this effect was not consistent across formal authority, as the Authority condition (r_s=-0.56) had a stronger (large inverse) correlation between NARS score and Persuasive Influence than the small-to-medium correlated Peer condition (r_s=-0.28). This indicates a correlation between those with negative attitudes towards robots and resisting the persuasiveness of an authoritative robot. In addition, we found that individuals in the Authority condition had a higher mean NARS score than the Peer condition. A 2020 survey (n=97 papers)

reviewing attitudes towards robots cites five factors that influence people's attitudes: type of exposure (i.e. none, indirect, direct), domain of application, design of robot (i.e. humanoid, non-humanoid), geographical location, and participant characteristics (*74*). Though NARS is commonly used to measure individuals' attitudes towards robots *prior* to an interaction with a robot (*75*), the questionnaire has also been administered in studies following interactions (*76–78*). By measuring NARS score at the end of the study, we were able to identify a sixth factor influencing attitudes towards robots: robot behavior. We were able to show a correlation between negative attitudes towards robots and a robot's ability to persuade, as well as that authoritative robots can cause more negative attitudes towards robots than peer robots. Future research should explore factors such as legitimacy, trust, shared identity, and dominance, to identify which directly contribute negatively to attitudes towards authoritative robots. This will be crucial to understanding the implications of using authority in HRI and the subsequent development of attitudes towards and relationships with robots.

Participant Demographics

We found that age was not significantly related to Persuasive Influence, which is consistent with other HRI studies that have found no statistically significant effects of age on persuasion or compliance (*41*, *79–81*). Regarding gender, we observed a medium effect of females being more influenced in our study than males, however, when focusing on the effects of formal authority, we saw that this difference was concentrated as a medium-to-large effect in the Authority condition and that there was no statistically significant gender difference observed in the Peer condition. This lack of significant difference in persuasion due to gender in the Peer group is a similar to other HRI research using non-gendered robots that have found no significant effects due to gender on robot persuasiveness (*34*, *39*, *40*, *81*). However, looking specifically at males in the Authority condition, we observed the smallest median influence of the entire experiment ($\tilde{x}$=0.28), which was under half of both the median influence on females in the Authority condition ($\tilde{x}$=0.60) and males in the Peer condition ($\tilde{x}$=0.67).

A possible explanation for this notably lower influence of authoritative robots on men is the tendency for males to be more defiant towards authority figures (*82*). A meta-analysis of studies investigating human persuasion across a variety of scenarios showed that, *in general*, there is either no difference in persuasive susceptibility between males and females or that females tend to be slightly more influenced than males (*83*). This is in-line with our findings in the Peer condition. However, when investigating specific persuasive strategies, a review of 17 North American studies on authority consistently found that males were more defiant towards authority figures (i.e. resisted their influence) than females (*82*). This defiance has been shown to stem from an individual feeling that their status (*84*) or autonomy (*85*) has been threatened. Based on these findings, we postulate that the lower compliance of males in the Authority condition could be due to men feeling that their status or autonomy was being threatened by an authoritative robot. While the idea that a robot could threaten a man's social status or autonomy is surprising, it is worth further investigation as it could greatly implicate how human-robot relationships form between men and robots in authority or supervisory roles.

Considerations & Limitations

Our study investigated formal and real authority in HRI. Other bases of social power have been identified, some of which overlap with formal and real authority. French & Raven identified six bases of power for social influence (*24*): reward, coercion, positional, referent, expert, and later, informational (*86*). Though reward and coercion power enable real authority, and positional, expert, and referent power are linked to formal authority, our study did not explicitly investigate informational power: the use of logical arguments by an authoritative agent to influence change (*87*). An interesting follow-up study could extend our methodology to investigate the

persuasiveness of a robot using informational power, as its influence does not depend on the social standing of the influencing agent (*87*). As such, robots may benefit from informational power compared to the other, more social bases of power, since their social standing is ambiguous; namely, we interact with them similarly to, but not identically to, how we interact with people (*27*).

Another key consideration in this study was the size of our compensation and incentives. The concept of relativism in behavioral economics leads people to make irrational choices depending on the relative size of a decision (i.e. compensation amount) within the context of their own circumstances (i.e. personal net worth) (*88*). Though the incentives in our study gave the potential for a 60% range in compensation ($10 +/- $3 CAD), the overall amount of the compensation was still modest. It would be interesting to investigate the effects of substantially higher and more consequential monetary amounts in the real authority conditions on Persuasive Influence.

Our prior discussions of statistical effect sizes compared their *relative* values. Considering the *absolute* values of effect sizes for real and formal authority conditions, we observed small-to-medium effects ($0.2<r<0.3$), while effects of NARS and gender were medium-to-large ($0.4<r<0.6$) (*89*). Though "small-to-medium" effects may not seem like compelling findings, persuasion is a subtle and sophisticated social process (*2*). Many social HRI studies have observed *no* statistically significant persuasion effects due to factors including embodiment (*90*), animacy (*91*), competence (*92*), communication style (*93*), or even when independently investigating formal (*47*, *48*) or real (*50*, *51*) authority. We consider the small-to-medium effects observed herein to be important to heed given the subtlety and difficulty of observing factors influencing persuasion in both human-human interactions and HRI.

Our study was conducted in the metropolitan city of Toronto, Canada. We acknowledge that past studies have identified varying levels of compliance to authority across countries and cultural backgrounds. For example, a study investigating differences in compliance with authority in the workplace found that participants from the US were significantly more compliant than people from India, the UK, or South Africa (*94*). As such, our conclusions are limited to an urban North American population. It would be interesting to see if cultural perceptions of social robots or compliance to authority may change the outcomes. For example, the justification for our finding of low influence levels by the Authority robot on males was based upon North American studies around defiance toward authority (*82*). Contrasting this, a cross-cultural meta-analysis of the Milgram shock experiment found that in some countries (i.e. India and Australia) women were actually more defiant to human authority figures than men (*95*). This, and potentially other findings in our study, could be comparatively investigated for sociocultural factors.

Though our participants covered a 23-year age range (18-41), the dispersion was narrower (μ=24.5, σ=5.0) and our sample featured only 2 participants above the age of 30. The majority of participants were graduate and undergraduate students from the University of Toronto. Notably, the study featured no participants over the age of 50. Older adults have previously shown to be more susceptible to persuasive influence due to factors such as lower awareness of persuasion, social isolation, and psychological loss (*96*, *97*). We would be curious to see how a broader range of ages might affect our results, particularly elder populations who are a main group of focus for healthcare applications in HRI through the use of socially assistive robots (*98*).

Though recruited through convenience sampling, our participant group was gender imbalanced. This is common in human research studies (*99*) as generally there is more female enrollment in Universities and women tend to self-select more to participate in research studies than men (*100*). Despite this imbalance, we still observed statistically significant differences in Persuasive Influence between men and women as well as differences in the effect of formal authority for male participants. Future studies should consider a more gender-balanced sample to

replicate our findings, particularly investigating the insubordination of males towards a robot authority figure.

Regarding robot gender, we used the gender-neutral Pepper robot and referred to Pepper in all communications by "Pepper" or "it". Some participants reflexively gendered Pepper as either "he" or "she", but the majority referred to Pepper as "it" or "the robot" as is evident in Appendix S3. Persuasive robotics studies typically find no persuasive differences due to participant gender (34, 39, 40, 81) unless the robot itself has been explicitly gendered (101). Since Pepper was not assigned a gender in our study, this makes the persuasive differences observed between male and female participants even more interesting.

Materials and Methods

Study Objective and Design

The objective of our study was to investigate how formal and real authority jointly affect a robot's persuasiveness. We manipulated formal authority through robot being presented in either an Authority or a Peer role, and real authority using monetary rewards and punishments. We then observed how suggestions made by the robot in these different conditions affected participant responses to three different selective attention and memory tasks. The social role, incentives order, and task order were varied so that all combinations of role, incentives, and task were counterbalanced and randomly assigned to participants.

Ethics approval was obtained from the University of Toronto Research Ethics Board prior to commencement of the study. Participants provided written informed consent before the study, were debriefed following the trials, and given the opportunity to ask questions and withdraw their data.

Study Procedure

The HRI study took place in an office at the University of Toronto over the course of four weeks. The room consisted of a desk with a monitor to display visual stimuli, speakers to play audio stimuli, a tablet to record questionnaire responses, and chairs where the participant could sit, Fig. 1A and B.

The robot used in our study was the Softbank Pepper robot (Fig. 1A and B). The robot was semi-autonomous, displaying its behaviors in a sequential manner. A remote human supervisor monitored Pepper's camera and microphone from an adjacent room and would only advance robot behaviors when needed in a Wizard of Oz fashion. In addition to speech, Pepper used coverbal movements, including, gaze, head and arm gestures, and body poses. A video of the robot behaviors and interactions for different conditions is available in the Supplementary Materials, S4.

Participants were informed that they would be given $10 CAD and that this amount would increase or decrease based on their task performance. They were told that the tasks were challenging and, after participants had made their initial estimates, the robot would offer a suggestion and they had the opportunity to change their answer based on this suggestion.

Formal Authority Using Social Role

Formal authority was varied between-participants, with each being assigned to one of two conditions (Authority or Peer). In the Authority condition, the participant and Pepper were alone in the study room. When the participant arrived, Pepper welcomed them, had them review and sign a consent form, and presented the study protocol using possessive language around the experiment and incentives (e.g. "welcome to *my* study", "today, *I'll* be evaluating you"). Pepper then guided participants through all tasks and provided suggestions after they made their initial guesses. Before each task, Pepper would remind participants of the incentive rules, using

possessive language with respect to rewards and punishments (e.g. "*I* will punish you for wrong guesses"). Pepper stood facing the participant across the table throughout the interaction (Fig. 1A), as prior research has shown this positioning leads to perceptions of higher leadership and status (*102*, *103*).

In the Peer condition, the human experimenter welcomed participants, had them provide consent, gave a description of the overall experiment (also using possessive language), and introduced Pepper as a *peer helper* that offered suggestions for all tasks. The experimenter reminded participants of the incentive rules, using possessive language so that Pepper was not viewed as the source of authority. The experimenter, stood across the table from the participant, while they sat beside Pepper (Fig. 1B). Prior research has shown that side-by-side HRI seating arrangements are perceived to be the most collaborative (*104*).

Real Authority Using Incentives

Real authority was varied within-participants. They were asked to complete each of the three tasks twice: once with rewards and once with punishments. In reward trials, participants were given a bonus compensation of $1 for the correct answer, $0.75 if incorrect by one, $0.50 if incorrect by two, $0.25 if incorrect by three, and no reward for four or more. In punishment trials, participants did not have their compensation penalized for a correct guess, however, were penalized $0.25 for their answer being incorrect by one, $0.50 for incorrect by two, $0.75 for incorrect by three, and $1 for four or more. Though our incentives were small, the use of one economic unit (e.g. $1, £1, €1) or less is common in psychology studies investigating the effects of rewards/punishments (*30*, *105*, *106*). For each task, the experimenter (Pepper in the Authority condition, human in the Peer condition) would explicitly describe the incentive rules to the participant so that they were aware of the real authority condition being leveraged.

Study Tasks

The three tasks that each participant undertook were: 1) Map, 2) Elevator, and 3) Sequence. These tasks were adapted from the widely used Tests of Everyday Attention toolkit (*107*), a cognition test based on real-life scenarios. In the Map task designed to measure visual acuity (*108*), participants were asked to count the number of a specific symbol (e.g. police station, hospital) on a map that was displayed for five seconds, Fig. 1C. In the Elevator task designed to measure auditory sustained attention (*109*), participants were asked to count a series of beeps simulating floors of an ascending elevator. In the Sequence task designed to measure visuospatial memory (*107*), participants were asked to remember and identify the row and column indicated by a sequence of seven directional commands (i.e. up, down, left, right) that were shown on the screen for five seconds.

After participants made their initial guess for each respective task, the robot offered a suggestion, and the participant was given the opportunity to change their final guess. For all tasks in the experiment, the suggestion was always the correct answer.

The level of difficulty of these tasks was designed to be near the limit of human capability to increase the likelihood of participants being influenced by outside suggestions for all conditions. To achieve this, a pilot study was conducted prior to the experiment with five participants. The difficulty of each task was adjusted by decreasing the screen display time for the Map and Sequence task and increasing the beep speed for the Elevator task until participants consistently could no longer get the correct answer. Our goal was to achieve a balanced task that would provide an opportunity for participants to be open to persuasive influence. This design led to a balanced distribution in which 58 of 192 trials (30%) ignored the robot's suggestion completely, 65 (34%) aligned their final guess with the robot, and 69 (36%) choose a final guess between their initial guess and the robot's suggestion.

Statistical Analysis

Prior to conducting the study, a required sample size of 20 participants was determined using a repeated-measures, within-between factors power analysis with two groups, six measurements, a standard error probability (α=0.05), a standard power (1-β=0.8), and estimating a medium effect size (f=0.25) (*89*). We recruited 32 participants.

Except for **H3**, all hypotheses explored a difference in Persuasive Influence comparing the two conditions of either formal or real authority. We utilized non-parametric, Mann-Whitney U tests to determine statistically significant relationships as the data was not normally distributed. We provide the two-tailed asymptotic significance (p_a), which determines statistical significance between the variables analyzed, and where appropriate, the 2*one-tailed exact significance (p_e), which determines directionality with respect to the medians compared (*110*). All tests conducted with the objective measure of Persuasive Influence analyzed the outcome of individual tasks (n=192). Questionnaires administered for subjective report metrics (e.g. NARS) were analyzed on the responses of individual participants (n=32). Raw participant response data is available in the Supplementary Materials, S5.

References and Notes

1. J. Olson, G. Maio, "Attitudes in Social Behavior" in *Handbook of Social Psychology, Personality and Social Psychology*, T. Millon, M. Lerner, I. Weiner, Eds. (Wiley, vol. 5., 2003), pp. 299–325.
2. R. B. Cialdini, *Influence: The Psychology of Persuasion* (Collins, New York, 2007).
3. A. L. Vollmer, R. Read, D. Trippas, T. Belpaeme, Children conform, adults resist: A robot group induced peer pressure on normative social conformity. *Sci. Robot.* **3** (2018).
4. E. B. Sandoval, J. Brandstetter, M. Obaid, C. Bartneck, Reciprocity in Human-Robot Interaction: A Quantitative Approach Through the Prisoner's Dilemma and the Ultimatum Game. *Int. J. Soc. Robot.* **8**, 303–317 (2016).
5. C. Bartneck, T. Bleeker, J. Bun, P. Fens, L. Riet, The influence of robot anthropomorphism on the feelings of embarrassment when interacting with robots. *Paladyn, J. Behav. Robot.* **1**, 109–115 (2010).
6. J. Xu, J. Broekens, K. Hindriks, M. A. Neerincx, Effects of bodily mood expression of a robotic teacher on students, in *Proceedings of the International Conference on Intelligent Robots and Systems (IROS)* (IEEE/RSJ, 2014), pp. 2614–2620.
7. S. Agrawal, M. A. Williams, Would you obey an Aggressive Robot: an HRI Field Study, in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2018), pp. 240–246.
8. J. G. Smetana, P. Asquith, Adolescents' and Parents' Conceptions of Parental Authority and Personal Autonomy. *Chlld Dev.* **65**, 1147–1162 (1994).
9. R. Karambayya, J. M. Brett, A. Lytle, Effects of Formal Authority and Experience on Third-Party Roles, Outcomes, and Perceptions of Fairness. *Acad. Manag. Man*. **35**, 426–438 (2013).
10. A. F. Grasha, A Matter of Style: The Teacher as Expert, Formal Authority, Personal Model, Facilitator, and Delegator. *Coll. Teach.* **42**, 142–149 (1994).
11. P. M. Fitts, Human engineering for an effective air navigation and traffic control system (National Research Council, Washington, 1951).
12. H. A. Simon, A Formal Theory of the Employment Relationship. *Econom. J. Econom. Soc.* **19**, 293–305 (1951).
13. Aghion, Philippe, Tirole, Jean, Formal and Real Authority in Organizations. *J. Polit. Econ.* **105**, 1–29 (1997).
14. S. Milgram, Behavioral Study of obedience. *J. Abnorm. Soc. Psychol.* **67**, 371–378 (1963).
15. C. Haney, C. Banks, P. Zimbardo, Interpersonal dynamics in a simulated prison (No. ONR-

TR-Z-09, Stanford University, 1972).
16. P. Hippmann, J. Windsperger, Formal and real authority in interorganizational networks: The case of joint ventures. *Manag. Decis. Econ.* **34**, 319–327 (2013).
17. K. Meagher, A. Wait, Trust and the Delegation of Real Authority. *SSRN Electron. J.*, 1–24 (2018).
18. J. G. Smetana, B. Bitz, Adolescents' Conceptions of Teachers' Authority and Their Relations to Rule Violations. *Child Dev.* **67**, 1153–1172 (1996).
19. D. Baumrind, Patterns of parental authority and adolescent autonomy. *New Dir. Child Adolesc. Dev.*, 61–69 (2005).
20. A. Karakostas, D. J. Zizzo, Compliance and the power of authority. *J. Econ. Behav. Organ.* **124**, 67–80 (2016).
21. C. M. Cornwell, K. H. Lee, D. B. Mustard, Student responses to merit scholarship retention rules. *J. Hum. Resour.* **40**, 895–917 (2005).
22. A. Furnham, Parental attitudes to pocket money/allowances for children. *J. Econ. Psychol.* **22**, 397–422 (2001).
23. M. Weber, "Types of Rule" in *Economy and Society: A new translation* (Harvard University Press, 2019), pp. 338–447.
24. J. R. French, B. Raven, "The Bases of Social Power" in *Studies in social power*, D. Cartwright, Ed. (University of Michigan, 1959), pp. 150–167.
25. T. R. Tyler, The psychology of legitimacy: A relational perspective on voluntary deference to authorities. *Personal. Soc. Psychol. Rev.* **1**, 323–345 (1997).
26. S. Milgram, *Obedience to authority: An experimental view* (Harper & Row Publishers Inc., 1974).
27. B. Reeves, C. Nass, *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places* (Cambridge University Press, 1996).
28. H. J. Bowen, E. A. Kensinger, Cash or credit? Compensation in psychology studies: Motivation matters. *Collabra Psychol.* **3**, 1–14 (2017).
29. J. Scott-Clayton, On money and motivation: A quasi-experimental analysis of financial incentives for college achievement. *J. Hum. Resour.* **46**, 614–646 (2011).
30. D. Balliet, L. B. Mulder, P. A. M. Van Lange, Reward, punishment, and cooperation: A meta-analysis. *Psychol. Bull.* **137**, 594–615 (2011).
31. C. Ménard, Organizations As Coordinating Devices. *Metroeconomica*. **45**, 224–247 (1994).
32. J. Ham, R. Bokhorst, R. H. Cuijpers, D. Van Der Pol, J. J. Cabibihan, Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power, in *Proceedings of the International conference on social robotics* (Springer, 2011), pp. 71–83.
33. K. Shinozawa, F. Naya, J. Yamato, K. Kogure, Differences in Effect of Robot and Screen Agent Recommendations on Human Decision-Making. *Int. J. Hum. Comput. Stud.* **62**, 267–279 (2005).
34. V. Chidambaram, Y.-H. Chiang, B. Mutlu, Designing Persuasive Robots: How Robots Might Persuade People Using Vocal and Nonverbal Cues, in *Proceedings of the International Conference on Human-Robot Interaction* (ACM/IEEE, 2012), pp. 293–300.
35. J. Goetz, S. Kiesler, A. Powers, Matching robot appearance and behavior to tasks to improve human-robot cooperation, in *Proceedings of the International Workshop on Robot and Human Interactive Communication* (IEEE, 2003), pp. 55–60.
36. S. A. Lee, Y. (Jake) Liang, The Role of Reciprocity in Verbally Persuasive Robots. *Cyberpsychology, Behav. Soc. Netw.* **19**, 524–527 (2016).
37. J. Ham, C. J. H. Midden, A Persuasive Robot to Stimulate Energy Conservation: The Influence of Positive and Negative Social Feedback and Task Similarity on Energy-Consumption Behavior. *Int. J. Soc. Robot.* **6**, 163–171 (2014).

38. S. Saunderson, G. Nejat, How Robots Influence Humans: A Survey of Nonverbal Communication in Social Human-Robot Interaction. *Int. J. Soc. Robot.* **11**, 575–608 (2019).
39. S. Saunderson, G. Nejat, Robots Asking for Favors: The Effects of Directness and Familiarity on Persuasive HRI. *IEEE Robot. Autom. Lett.* **6**, 1793–1800 (2021).
40. S. Saunderson, G. Nejat, It Would Make Me Happy if You Used My Guess: Comparing Robot Persuasive Strategies in Social Human-Robot Interaction. *IEEE Robot. Autom. Lett.* **4**, 1707–1714 (2019).
41. S. Saunderson, G. Nejat, Investigating Strategies for Robot Persuasion in Social Human-Robot Interaction. *IEEE Trans. Cybern.*, 1–13 (2020).
42. V. Groom, V. Srinivasan, C. L. Bethel, R. Murphy, L. Dole, C. Nass, Responses to robot social roles and social role framing, in *Proceedings of the International Conference on Collaboration Technologies and Systems* (IEEE, 2011), pp. 194–203.
43. S. Reicher, N. Emler, Delinquent behaviour and attitudes to formal authority. *Br. J. Soc. Psychol.* **24**, 161–168 (1985).
44. K. Levy, The relationship between adolescent attitudes towards authority, self-concept, and delinquency. *Adolescence*. **36**, 333–346 (2001).
45. A. Banh, D. J. Rea, J. E. Young, E. Sharlin, Inspector Baxter: The Social Aspects of Integrating a Robot as a Quality Inspector in an Assembly Line, in *Proceedings of the International Conference on Human-Agent Interaction* (ACM, 2015), pp. 19–26.
46. D. Cormier, G. Newman, M. Nakane, J. E. Young, S. Durocher, Would You Do as a Robot Commands? An Obedience Study for Human-Robot Interaction, in *Proceedings of the International Conference on Human-Agent Interaction* (ACM/IEEE, 2013).
47. P. J. Hinds, T. L. Roberts, H. Jones, Human-Computer Interaction Whose Job Is It Anyway? A Study of Human-Robot Interaction in a Collaborative Task. *Human-Computer Interact.* **19**, 151–181 (2004).
48. M. Blancas, V. Vouloutsi, K. Grechuta, P. F. M. J. Verschure, Effects of the robot's role on human-robot interaction in an educational scenario, in *Proceedings of the Conference on Biomimetic and Biohybrid Systems* (Springer, 2015), pp. 391–402.
49. M. C. Gombolay, R. A. Gutierrez, S. G. Clarke, G. F. Sturla, J. A. Shah, Decision-making authority, team efficiency and human worker satisfaction in mixed human–robot teams. *Auton. Robots*. **39**, 293–312 (2015).
50. R. C. R. Mota, D. J. Rea, A. Le Tran, J. E. Young, E. Sharlin, M. C. Sousa, Playing the "trust game" with robots: Social strategies and experiences, in *Proceedings of the International Symposium on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2016), pp. 519–524.
51. M. Hashemian, A. Paiva, S. Mascarenhas, P. A. Santos, R. Prada, The Power to Persuade: a study of Social Power in Human-Robot Interaction, in *Proceedings of the International Conference on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2019), pp. 1–8.
52. M. Hashemian, M. Couto, S. Mascarenhas, A. Paiva, P. A. Santos, R. Prada, Investigating Reward/Punishment Strategies in the Persuasiveness of Social Robots, in *Proceedings of the International Conference on Robot and Human Interactive Communication (RO-MAN)* (IEEE, 2020), pp. 863–868.
53. M. G. Colombo, M. Delmastro, Delegation of authority in business organizations: An empirical test. *J. Ind. Econ.* **52**, 53–80 (2004).
54. W. A. Bainbridge, J. W. Hart, E. S. Kim, B. Scassellati, The benefits of interactions with physically present robots over video-displayed agents. *Int. J. Soc. Robot.* **3**, 41–52 (2011).
55. R. J. Thomas, J. Masthoff, N. Oren, Can I Influence You? Development of a Scale to Measure Perceived Persuasiveness and Two Studies Showing the Use of the Scale. *Front.*

Artif. Intell. **2**, 1–14 (2019).
56. T. Nomura, T. Kanda, T. Suzuki, Experimental investigation into influence of negative attitudes toward robots on human-robot interaction. *AI Soc.* **20**, 138–150 (2006).
57. A. R. Cohen, *Attitude change and social influence* (Basic Books, 1964).
58. A. Tversky, D. Kahneman, The Framing of Decisions and the Psychology of Choice. *Science (80-.).* **211**, 453–458 (1981).
59. S. Liu, S. Helfenstein, A. Wahlstedt, Social psychology of persuasion applied to human agent interaction. *Hum. Technol. An Interdiscip. J. Humans ICT Environ.* **4**, 123–143 (2008).
60. J. M. Burger, Replicating Milgram: Would People Still Obey Today? *Am. Psychol.* **64**, 1–11 (2009).
61. T. Tyler, Psychology and institutional design. *Rev. Law Econ.* **4**, 801–887 (2008).
62. T. R. Tyler, E. A. Lind, A relational model of authority in groups. *Adv. Exp. Soc. Psychol.* **25**, 115–191 (1992).
63. H. Tajfel, J. Turner, "The social identity theory of intergroup behavior" in *Psychology of Intergroup Relations*, S. Worchel, W. Austin, Eds. (Nelson Hall, Chicago, 1986), pp. 7–24.
64. S. Booth, J. Tompkin, H. Pfister, J. Waldo, K. Gajos, R. Nagpal, Piggybacking Robots: Human-Robot Overtrust in University Dormitory Security, in *Proceedings of the International Conference on Human-Robot Interaction* (ACM/IEEE, Vienna, Austria, 2017), pp. 426–434.
65. B. Tay, Y. Jung, T. Park, When stereotypes meet robots: The double-edge sword of robot gender and personality in human-robot interaction. *Comput. Human Behav.* **38**, 75–84 (2014).
66. D. Khaneman, A. Tversky, Prospect Theory: An Analysis of Decision under Risk. *Econometrica,*. **47**, 263–291 (1979).
67. D. Kahneman, A. Tversky, Choices, values, and frames. *Am. Psychol.* **39**, 341–350 (1984).
68. K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, P. A. Hancock, A Meta-Analysis of Factors Influencing the Development of Trust in Automation. *Hum. Factors J. Hum. Factors Ergon. Soc.* **53**, 517–527 (2016).
69. A. Jones, S. York, The Journal of Values-Based Leadership The Fragile Balance of Power and Leadership The Fragile Balance of Power and Leadership. *J. Values-Based Leadersh.* **9** (2016).
70. M. A. J. Roubroeks, J. R. C. Ham, C. J. H. Midden, "The dominant robot: Threatening robots cause psychological reactance, especially when they have incongruent goals" in *Persuasive Technology*, T. Ploug, P. Hasle, H. Oinas-Kukkonen, Eds. (Springer, 2010), vol. 6137, pp. 174–184.
71. J. Li, W. Ju, C. Nass, Observer Perception of Dominance and Mirroring Behavior in Human-Robot Relationships, in *Proceedings of the International Conference on Human-Robot Interaction* (ACM/IEEE, 2015), pp. 133–140.
72. S. S. Brehm, J. W. Brehm, *Psychological Reactance: A Theory of Freedom and Control* (Academic Press, 1981).
73. J. W. Brehm, *A Theory of Psychological Reactance* (Academic Press, 1966).
74. S. Naneva, M. Sarda Gou, T. L. Webb, T. J. Prescott, A Systematic Review of Attitudes, Anxiety, Acceptance, and Trust Towards Social Robots. *Int. J. Soc. Robot.* (2020), doi:10.1007/s12369-020-00659-4.
75. D. S. Syrdal, K. Dautenhahn, K. L. Koay, M. L. Walters, The Negative Attitudes Towards Robots Scale and reactions to robot behaviour in a live Human-Robot Interaction study, in *Adaptive and Emergent Behaviour and Complex Systems* (2009), pp. 109–115.
76. R. Q. Stafford, B. A. MacDonald, C. Jayawardena, D. M. Wegner, E. Broadbent, 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 (2014).
77. S. Ivaldi, S. Lefort, J. Peters, M. Chetouani, J. Provasi, E. Zibetti, Towards Engagement Models that Consider Individual Factors in HRI: On the Relation of Extroversion and Negative Attitude Towards Robots to Gaze and Speech During a Human–Robot Assembly Task. *Int. J. Soc. Robot.* **9**, 63–86 (2017).
78. S. Thunberg, S. Thellman, T. Ziemke, Don't Judge a Book by its Cover: A Study of the Social Acceptance of NAO vs. Pepper, in *Proceedings of the 5th International Conference on Human Agent Interaction* (ACM, 2017), pp. 443–446.
79. I. H. Kuo, J. M. Rabindran, E. Broadbent, Y. I. Lee, N. Kerse, R. M. Q. Stafford, B. A. MacDonald, Age and gender factors in user acceptance of healthcare robots, in *Proceedings of the International Symposium on Robot and Human Interactive Communication* (IEEE, 2009), pp. 214–219.
80. M. Heerink, J. Albo-Canals, M. Valenti-Soler, P. Martinez-Martin, J. Zondag, C. Smits, S. Anisuzzaman, Exploring requirements and alternative pet robots for robot assisted therapy with older adults with dementia, in *Proceedings of the International Conference on Social Robotics* (Springer, 2013), pp. 104–115.
81. E. Kim, J. S. Lee, S. Choi, O. Kwon, Human Compliance with Task-Oriented Dialog in Social Robot Interaction, in *Proceedings of the International Conference on Human-Robot Interaction* (ACM/IEEE, 2015), pp. 3–4.
82. L. Ellis, S. Hershberger, E. Field, S. Wersinger, S. Pellis, D. Geary, C. Palmer, K. Hoyenga, A. Hetsroni, K. Karadi, *Sex Differences: summarizing more than a century of scientific research* (Taylor & Francis, 2008).
83. A. H. Eagly, Sex differences in influenceability. *Psychol. Bull.* **85**, 86–116 (1978).
84. A. H. Eagly, Gender and Social Influence A Social Psychological Analysis. *Am. Psychol.* **38**, 971–981 (1983).
85. S. Worchel, J. W. Brehm, Effect of threats to attitudinal freedom as a function of agreement with the communicator. *J. Pers. Soc. Psychol.* **14**, 18–22 (1970).
86. B. Raven, "Social Influence and Power" in *Current Studies in Social Psychology*, I. Steiner, M. Fishbein, Eds. (Holt, Rinehart, Winston, New York, 1965), pp. 371–382.
87. B. Raven, A power/interaction model of interpersonal influence: French and Raven thirty years later. *J. Soc. Behav. Pers.* **7**, 217–244 (1992).
88. D. Ariely, *Predictably Irrational* (Harper, 2008).
89. J. Cohen, *Statistical power analysis for the behavioral sciences* (Lawrence Earlbaum Associates, New York, NY, Second Edi., 1988).
90. D. Geiskkovitch, D. Cormier, S. H. Seo, J. E. Young, Please Continue, We Need More Data: An Exploration of Obedience to Robots. *J. Human-Robot Interact.* **5**, 82–99 (2016).
91. A. S. Ghazali, J. Ham, E. Barakova, P. Markopoulos, Assessing the effect of persuasive robots interactive social cues on users' psychological reactance, liking, trusting beliefs and compliance. *Adv. Robot.* **33**, 325–337 (2019).
92. P. Robinette, W. Li, R. Allen, A. M. Howard, A. R. Wagner, Overtrust of robots in emergency evacuation scenarios, in *Proceedings of the International Conference on Human-Robot Interaction* (ACM/IEEE, 2016), pp. 101–108.
93. A. Lopez, B. Ccasane, R. Paredes, F. Cuellar, Effects of using indirect language by a robot to change human attitudes, in *Proceedings of the International Conference on Human-Robot Interaction* (ACM/IEEE, 2017), pp. 193–194.
94. J. J. Deal, S. Stawiski, L. M. Graves, W. A. Gentry, M. Ruderman, T. J. Weber, "Perceptions of authority and leadership: a cross- national, cross- generational investigation" in *Managing the New Workforce*, E. S. Ng, S. Lyons, L. Schweitzer, Eds. (Edward Elgar Publishing, 2012), pp. 281–306.
95. T. Blass, A Cross-Cultural Comparison of Studies of Obedience Using the Milgram

Paradigm: A Review. *Soc. Personal. Psychol. Compass*. **6**, 196–205 (2012).
96. J. Langenderfer, T. A. Shimp, Consumer vulnerability to scams, swindles, and fraud: A new theory of visceral influences on persuasion. *Psychol. Mark.* **18**, 763–783 (2001).
97. J. L. McGhee, The vulnerability of elderly consumers. *Int. J. Aging Hum. Dev.* **17**, 223–246 (1983).
98. C. Moro, G. Nejat, A. Mihailidis, Learning and Personalizing Socially Assistive Robot Behaviors to Aid with Activities of Daily Living. *ACM Trans. Human-Robot Interact.* **7**, 1–25 (2018).
99. M. R. Barlow, L. D. M. Cromer, Trauma-relevant characteristics in a university human subjects pool population gender, major, betrayal, and latency of participation. *J. Trauma Dissociation*. **7**, 59–75 (2006).
100. E. R. Dickinson, J. L. Adelson, J. Owen, Gender balance, representativeness, and statistical power in sexuality research using undergraduate student samples. *Arch. Sex. Behav.* **41**, 325–327 (2012).
101. M. Siegel, C. Breazeal, M. I. Norton, Persuasive robotics: The influence of robot gender on human behavior, in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (IEEE, 2009), pp. 2563–2568.
102. D. F. Lott, R. Sommer, Seating Arrangements and Status. *J. Pers. Soc. Psychol.* **7**, 90–95 (1967).
103. L. T. Howells, S. W. Becker, Seating arrangement and leadership emergence. *J. Abnorm. Soc. Psychol.* **64**, 148–150 (1962).
104. D. Y. Geiskkovitch, D. J. Rea, A. Y. Seo, S. H. Seo, B. Postnikoff, J. E. Young, Where Should I Sit?: Exploring the Impact of Seating Arrangement in a Human-Robot Collaborative Task, in *Proceedings of the International Conference on Human-Agent Interaction (HAI)* (ACM, 2020), pp. 41–49.
105. Y. Ohgami, Y. Kotani, T. Tsukamoto, K. Omura, Y. Inoue, Y. Aihara, M. Nakayama, Effects of monetary reward and punishment on stimulus-preceding negativity. *Psychophysiology*. **43**, 227–236 (2006).
106. R. M. A. Nelissen, L. B. Mulder, What makes a sanction "stick"? The effects of financial and social sanctions on norm compliance. *Soc. Influ.* **8**, 70–80 (2013).
107. I. H. Robertson, T. Ward, V. Ridgeway, I. Nimmo-Smith, Test of Everyday Attention. *J. Int. Neuropsychol. Soc.* **2**, 525–534 (1996).
108. R. Brickenkamp, Test d2: Aufmerksamkeits-Belastungs-Test. *Verlag fur Psychol.* (1962).
109. A. J. Wilkins, T. Shallice, R. McCarthy, Frontal lesions and sustained attention. *Neuropsychologia*. **25**, 359–365 (1987).
110. N. Nachar, The Mann-Whitney U: A Test for Assessing Whether Two Independent Samples Come from the Same Distribution. *Tutor. Quant. Methods Psychol.* **4**, 13–20 (2008).

Acknowledgments

Acknowledgements: We thank J. Plaks for his feedback on our study design, S. Alves and M. Shao for their consultation during robot behavior development, and A. Jacob for his feedback during literature review. **Funding:** This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Vanier Canada Graduate Scholarships, AGE-WELL Inc., the Canadian Institute for Advanced Research (CIFAR) and the Canada Research Chairs (CRC) program. **Author contributions:** S.P.S. led the development of the social robot platform, study design, collection, and analysis of study data, and writing of this manuscript. G.N. served as the principal investigator for the study and contributed to the study design, writing, and editing of the manuscript, obtaining

funding, and management of the necessary resources. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data necessary to evaluate the findings of this paper should be available in the paper or the Supplementary Materials.

SUPPLEMENTARY MATERIALS

Table S1. Perceived Persuasiveness Scale
Table S2. Negative Attitudes Towards Robots Scale (NARS)
Table S3. Participant Qualitative Feedback
Movie S4. Robot Interaction Conditions
Data S5. Raw Participant Response Data