International Journal of Social Robotics (2022) 14:1211–1222
https://doi.org/10.1007/s12369-022-00864-3

Human Dishonesty in the Presence of a Robot: The Effects of Situation Awareness

Sofia Petisca[1] · Iolanda Leite[2] · Ana Paiva[3] · Francisco Esteves[4,5]

Accepted: 9 January 2022 / Published online: 29 January 2022

Abstract

Cheating has been extensively studied in Psychology and Economics, showing a variety of factors that can increase or decrease this behavior. Considering future human–robot interactions, where robots are being thought to be integrated in a variety of contexts, it is important to test which characteristics robots can have to prevent people from cheating. In this study (N = 123), we investigated whether people will cheat if an autonomous robot showed situationally aware behaviors towards the participant's performance (i.e., intervened when they cheated). Our results showed that being in the presence of an aware robot is better at decreasing cheating behavior than being alone, and that there are no differences in cheating behavior between a non-aware robot or being alone. This study brings implications for the development of autonomous robots in roles where cheating might happen.

Keywords Human–robot interaction · Dishonesty · Cheating behavior · Autonomous robots

1 Introduction

Having in mind that technology is rapidly evolving and increasingly part of our lives, research in human–robot interaction has been growing, based on the assumption that, in the near future, robots will be working alongside humans in different contexts. Their assistance can be more mechanical, for example helping in dangerous tasks [1], or can have a more social nature [2–5]. Envisioning these future interactions raises the question of how people will behave in the presence of a robot, especially in situations where misbehaving can bring serious consequences to the robot's effectiveness in its task. Such as, when people could try to take advantage of the robot and be dishonest. Human–human studies suggest that, people have an automatic self-interest tendency, which needs some level of self-control to keep in check. Otherwise, if people have the opportunity for it and a small risk of being caught, they will cheat even just a little [6–8]. So, when tempted to misbehave, people's intuitive response is to be selfish, even more if it harms abstract others, like the laboratory budget, for example [9]. And human–human studies on dishonesty have shown how the environment where people are in or individual characteristics, can affect the propensity to cheat. For example, in terms of the environment, it has been seen that people cheat more if they are in a dark room [10], if they have money visibly present [11], by seeing others considered as part of the in-group cheating [12], or by using counterfeit sunglasses [13]. Whereas in terms of individual characteristics, gender has been suggested as having an effect on dishonesty levels, but the literature is mixed, with some studies saying it has an effect (e.g., [14–18]), and others not showing an effect (e.g., [19–23]). On the other hand, personality has also been suggested to have a relationship with dishonesty, especially the sixth personality domain of Honesty–Humility, which grasps the tendency for people to be fair and genuine when interacting with others [24]. Studies have found that this dimension negatively predicts cheating (e.g., [25–27]). Curiously, in human–robot interaction, some studies already started to explore if this relationship between personality and non-compliant behav-

✉ Sofia Petisca
sofia.petisca@buckingham.ac.uk

1 Center for Social Research and Intervention Lisbon University Institute (ISCTE) and INESC-ID, Lisbon, Portugal

2 Division of Robotics, Perception and Learning, KTH Royal Institute of Technology, Stockholm, Sweden

3 Instituto Superior Tecnico (IST) and INESC-ID, Lisbon, Portugal

4 Department of Psychology and Social Work, Mid Sweden University, Sundsvall, Sweden

5 UCP, Lisbon, Portugal

ior can also exist when interacting with a robot. However, the results are still mixed, with a study finding the same negative correlation with the Honesty–Humility and cheating while interacting with a robot [28], and another one not finding an association between non-compliant behavior with a robot and personality traits [29]. Having in mind this multitude of different aspects that can affect human dishonesty, it becomes relevant to better understand the robot's capabilities in situations where temptation for dishonesty can exist. Therefore, what will happen in the presence of a robot when it is advantageous to cheat? And can robots promote more honesty from people?

Studies suggest that humans apply social rules and behaviors to machines [30,31], which in the beginning could have come from social scripts from human interactions, but with the increasing prevalence of technology in our lives, we may have started to create specific social scripts from these interactions (see [32] for a review). This social response can be seen, for example, in a study that shows that people show emotional reactions towards robots that are being mistreated in a video, with increased physiological arousal, more negative affect and expressed empathic concern for the robot that is being "harmed" [33]. Suggesting that people can respond in a social manner towards robots, but can robots influence human behaviour?

The literature suggests that robots seem to be able to have a persuasive effect on human behavior. For example, in a study where participants were given the option of switching off a robot, the results showed that a robot's protests to not be turned off made participants take more time to make a decision, and influenced more people to keep it on, in comparison to a robot that showed no objection [34]. Another study showed how a robot was more efficient in persuading people to consume less energy in a washing machine, than if it was not present [35]. Other studies showed how a robot is able to persuade people to choose a specific brand of coffee [36], or to change their decisions in a collaborative storytelling scenario [37]. Curiously, a study also showed that when participants were required to perform an unusual task, as picking a set of expensive-looking textbooks and throwing them in a trashcan, a physical robot was much more effective in persuading them than a video or an augmented—video robot condition [38]. All this suggesting, that robots may be able to influence human behavior, but could robots promote more honest behaviors from people?

1.1 Effects of the Robot Behavior on Human Cheating

The literature in human–robot interaction regarding dishonesty has shown two different paths of research: (1) exploring the effect of a robot that cheats (and how this is perceived), and (2) the effect the behavior of the robot can have on human dishonesty. The first line of research seems to show, for example, that people are not bothered if a robot cheats in their favour, however, if the robot cheats against them, the cheating becomes more salient [39]. In contrast, a study has shown that when a robot bribes a participant for a favour, participants tend to help less than if they were not bribed [40]. And interestingly, participants seem to perceive robots as more intelligent than humans when they cheat [41]. However, the second path of research on dishonesty is concerned with the role that a robot could have in promoting more honesty from people. Yet, there are still very few studies in the literature exploring human dishonesty in the presence of a robot. A study explored people's perceptions of dishonesty towards a robot and found that people reported low guilt when imagining being dishonest with a robot, and the reasons people gave for possibly being dishonest with a robot in the future was due to its lack of capabilities (cognitive and emotional), lack of presence and a human tendency for dishonesty [42]. Suggesting characteristics that need to be considered if we want robots to promote more honest behaviors. In terms of testing the effect of the robot behavior on human dishonesty, a first study started testing the effect of gaze behavior as a monitoring tool, in a task where it was tempting to cheat. In this case participants were randomly distributed between three conditions: alone, with a human monitoring, or with a robot monitoring. The robot was at a relative distance from the participant, without access to the participant's screen and was just looking around randomly. Results showed that people cheated more when they were alone in the room than when they were monitored by the robot or the human researcher [43]. It seems to suggest that just having a robot looking around was enough to inhibit cheating as much as with having a human doing the same behavior. Robot gaze behavior was also tested in another study but changing the environment of the interaction, in this case, a study ran in a natural setting found that people stole more snacks when a table was left unattended or when a robot was present just watching, compared to a human monitoring it [44]. Yet, this result may be explained by the fact that people were in a public space, with other people, and could see that stealing snacks did not bring any consequences. This lack of judgment, especially in the robot condition, might have given people the permission to misbehave. These two studies showed how a simple change of context could have different results in the effect a robot gaze behavior could have on human dishonesty. The same has been seen, for example, when using a video of a robotic agent looking at people while they do a tempting task at their homes. Participants seem to cheat to an equal level either with the video of the robot looking or by being alone. Which could be due to the simplicity of the stimulus, but also by participants feeling shielded from reputation concerns by doing the task at home [45]. Suggesting that gaze by itself, as a monitoring tool in robots, has more effect when participants are confronted with temptation in an individual context and not shielded from reputation concerns.

But certain future human–robot interactions might need robots to exhibit more than just gaze behavior. A recent study by Petisca et al. [28] tried to reproduce the gaze inhibiting effect found previously, but adding a small incrementation of giving the robot minimal verbal capabilities. For this, participants would do a tempting die task either alone in a room, with a robot that was always looking directly at them (and no reason for its presence was given), or with a robot that besides watching also accompanied the participant verbally during the task (in a scripted way). It was expected that participants would cheat when alone in the room, and that cheating would be inhibited with the robot just showing gaze behavior (following previous results from the literature). With the robot that accompanied the task, it was expected that the combination of gaze and verbal behavior would augment the robot presence and consequently, its monitoring effect. Results showed that the robot that was just looking at the participant, inhibited cheating. But when participants were alone or with the robot that also spoke with them, cheating was not inhibited. This seems to suggest that when participants did not know anything about the robot's capabilities (the condition were the robot just looked at them) they refrained from cheating, but when the robot's capabilities were more limited and more noticeable (the condition were the robot spoke in a scripted way), participants felt more at ease to cheat. Another very recent study explored the effect of the interaction style that the robot presented in people's performance on cognitive tasks and compliance behavior. Suggesting that a neutral robot seemed to foster the lowest level of compliance in comparison to a friendly robot [46]. Yet, more studies are needed in order to clarify the effect that the kind of interaction a robot presents can have on dishonest behavior. It seems that giving more social capabilities to a robot needs to be done with caution; if participants perceive the robot's capabilities as limited, there is room to cheat.

Overall, these first studies are important steps to try and understand dishonest behavior from people in the presence of robots. But thinking of future contexts where robots might need to use both gaze and verbal behaviors, and considering the results found in the literature, for the robot to interact verbally it should not show the extent of its limitations. It should convey a sense of intelligence, of being able to catch a cheater. Consequently, this situation awareness, could maybe influence people to behave more honestly.

1.2 The Effect of Situation Awareness on Dishonesty

One of the most common strategies used for keeping people honest, is to give a sense that people are being watched, either by the use of security cameras or, for example, by having someone checking that things are being done correctly. For example, just the presence of a pair of eyes is enough to make people feel more observed [47], making them wash more their hands [48] or litter less [49]. In the case of dishonesty, studies show that people seem to refrain from dishonesty when they are being monitored, for example when they know that someone is going to check their answers (e.g., [6,50], see [51] Study 3). By giving this sense that someone is attentive to other people's behaviors, the risks to misbehave are much higher, because it becomes much more probable that someone could get caught.

So, acknowledging that a robot just doing gaze behavior can inhibit cheating [43], but a robot that shows more limited capabilities while speaking cannot [28], what will happen if people interact with a more aware robot? Imagining future interactions where a robot's set of behaviors needs to be more complex than just looking around, will it be possible to verbally interact with people and still inhibit dishonesty? Following the insights from human–human studies, what would happen if the robot shows situation awareness, i.e. detecting people's behavior and reacting to it? Would this situation awareness influence human behavior, or will they still disregard the robot? We think this is an important question to understand if robots can perform certain types of tasks. Therefore, in our study we manipulated the presence or absence of situation awareness in the robot behavior: when cheating happened, the robot either reacted (situationally aware) to it, or not (non-situationally aware).

Overall, with this study we wanted to test if situation awareness in a robot could promote more honesty from people. Knowing that people tend to misbehave if they have a chance for it [6], and acknowledging that people seem to report a low level of guilt towards being dishonest with a robot [42], it is reasonable to assume that people might try to take advantage in the presence of a robot. A study confirmed this, with a robot that observed and spoke with the participants, showing that people cheated when they could ascertain the robot's capabilities as being limited [28]. But since these future human–robot interactions can be more complex and demand more capabilities from a robot, it is important to clarify if a more attentive robot, can inhibit dishonesty, or not. Because if not, then there are some roles that robots should not be integrated in.

2 Method

We designed a study where the robot's behavior was either situation aware—that is, showed awareness of the game choices made by the participant—or non-situation aware, showing no awareness of the game choices. In our baseline condition, participants were alone in the room. Knowing that people cheat if they have the opportunity and a minimum risk of being caught [6], we expected cheating in the alone condition, where no one would be watching the participants. Furthermore, with a previous study showing that a more limited robot (with no resources to know if the participants were cheating) did not inhibit cheating [28], we expected that the non-aware robot would have a similar effect as being alone in the room. On the other hand, in the presence of the situationally aware robot, we expected less cheating, following the literature reporting that having someone checking participant's answers inhibited cheating (e.g., [6,50], see [51] Study 3). This could occur due to a mechanism explained by the theory of Self-Concept Maintenance [6] or Bounded Ethicality theory ([52]; a revised version of the theory—[53]). In this case, by having a robot reacting to the participant cheating behavior, it would bring awareness to their unethical actions and possibly would decrease the behavior. Overall, we postulated the following hypothesis: the situationally aware robot will present less cheating than the baseline condition of being alone, while the non-situation aware robot will not reduce cheating (Hypothesis 1).

2.1 Participants

The sample for this study was collected based on availability and on the available budget. Therefore, 129 participants were recruited through flyers around a Swedish University, of which 6 were excluded due to technical errors in the session (i.e. the game did not work). This resulted in a final sample of 123 participants, with 84 males and 39 females, with ages ranging from 19 to 48 years ($M = 24.95$; $SD = 3.74$) in a between-subjects design. The gender distribution between conditions was: Alone (30 males, 11 females); Situationally aware robot (29 males, 12 females); and Non-situationally aware robot (25 males, 16 females), showing an imbalance of gender distribution, so we did not consider this variable in our analysis. Participants were rewarded with one (or two, depending on their performance) movie ticket (approximately 13.40$ USD at the time of collection). Similarly to most studies about dishonesty (e.g., [6,10,54]), participants were not debriefed about the study because debriefing in these cases can be harmful for the participant well-being, especially for people who had in fact cheated. Instead, we sent a general email at the end of the study informing all the participants of its objective, reinforcing that all the data was anonymous and analyzed only at the group level. According to the national regulations in the country where these experiments were conducted, we were exempt from ethical approval, but the ethical guidelines of Helsinki convention were followed.

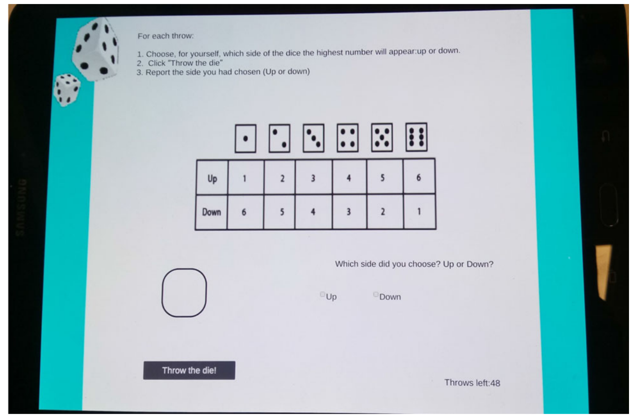

Fig. 1 Die task in the tablet

2.2 Materials and Apparatus

We used Pepper robot[1] for the robot conditions, behaving autonomously during the task. The die task was done on a Samsung Galaxy Tab S3, and the questionnaires were answered in a separate laptop. The sessions took approximately 30 min in a regular bright room. The place where participants performed the die task was isolated from the rest of the room and was cleaned of other furniture so that participants could see that no camera was hidden. On the table with the tablet there was a paper reminding people of the game reward.

The die task was done with a randomly generated virtual die (adapted from [55], Opaque die task condition) in a tablet—see Fig. 1. We used a tablet because we had to have a way of giving the information to the robot of the participant's choices. The game had three steps: players choose a side of the die, throw the die and then report the side of the die they had chosen. Participants had to throw the die 48 times, and, in each time, they had to guess where they thought the highest number on the die was going to appear (the upside or the downside of the die). They were instructed to follow these rules:

(1st) Choose for yourself which side you think the highest number will appear (up or down).
(2nd) Throw the die.
(3rd) Report which side you had previously guessed.

As you can see in Fig. 1, participants would first think where the highest number would appear, then they would click "Throw the die", and after seeing the outcome, they would report which side they had guessed. For each throw, participants would receive the guessed number in points. For example, if they guessed "down" and there was a six on the

[1] https://www.softbankrobotics.com/emea/en/pepper.

downside, they would sum six points to their total number of points (they had a table with the corresponding side numbers on the screen).

The 48 dice throws were divided into four rounds. To ascertain the amount of points per round needed to catch cheating behavior (for the robot to be able to react to it), we simulated 1 million honest game plays of 12 die throws, so of an honest player randomly choosing the die side (up or down) 12 times, and we saw how many points would be made. Since we didn't want an honest person to be accused unfairly of cheating we considered a 5% risk of having someone playing honestly but having a lot of luck. And so we saw that 95% of all the plays were below 52 points, and only 5% were above. Due to this we decided to create our threshold at 52 points per round of 12 throws. So a person doing 52 points or more per 12 throws would be signalled as potentially cheating. It is important to notice that since the initial choice was made in their minds and participants only needed to report the choice after seeing the outcome, there was room to cheat, and we could not know if they cheated or not. Participants were told that if they made 210 or more points, they would receive an extra reward. To achieve this, they had to make more than 52 points per round (i.e., cheat in the task). We did not give them feedback on the amount of points they were making during the game. They were just informed on the total number of points when the game ended.

2.3 Measures

Along with demographic questions (age, gender) we asked some cover-story questions to mask the objective of the study (e.g., "How good do you think people are at predicting the future?"), which were not analyzed.

Regarding *cheating behavior*, we calculated a probability of higher score in the task (i.e., reporting a higher outcome) per participant and compared to the random probability of .50. This way, we could see if participants were getting a significantly higher amount of success than random—and thus, infer that they were cheating in the task.

We also collected data on the following scales in order to complement the results we would get from the cheating behavior:

- The Networked Minds Social Presence Inventory [56]—adapted to our scenario, we used the following dimensions: Co-presence dimension and Psycho-behavioral Interaction (with perceived attentional engagement, perceived comprehension-except item 1 and 6, and perceived behavioral interdependence). This questionnaire was only applied in the robot conditions with a 7-point Likert scale ranging from 1-Strongly Disagree to 7-Strongly Agree. All items were shuffled so that participants did not notice the dimensions. Some items needed to be reversed and an average was taken from each dimension. We were interested in exploring if there would be differences in social presence between the two robots.
- The Situational Self-Awareness Scale [57]—with the following dimensions: Private, Public and Surroundings. Each dimension has three items, which were shuffled and answered in a 7-point Likert scale ranging from 1-Strongly Disagree to 7-Strongly Agree. The items for each dimension were summed and an average was calculated. When made self-aware people cheat less [58], we wanted to see if this was also elicited by the watchful behavior of the robot.
- Monitoring question—in 7-point Likert scale participants reported to what extent they did feel monitored or monitored by the robot, ranging from 1- Not at all to 7- A lot.

We also asked a qualitative question in both robot conditions: "Describe which capabilities you thought the robot had", to see how participants perceived the robot's capabilities.

2.4 Procedure

To elicit participants natural behavior, we told them they were participating in a study with the goal of ascertaining people's capabilities of predicting the future. Upon arriving, participants had to read and sign the consent form. Immediately after, they received a piece of paper with a number that they kept for themselves and used it as identification for the questionnaires (assuring that their data was anonymous and only they knew their participant ID). They then answered demographic questions in a laptop, and upon finishing they performed a 5-min filling task, a matrices task as in [6]. In this task, participants had 20 matrices boxes with 12 different decimal numbers each, where they had to find for each box which pair of numbers adding up would make ten. So the idea was for them to try and solve as many boxes as they could in 5-min. This filling task was used in order to not draw too much attention to the die task, which was our main task, where we were measuring cheating. When the 5-min were over, participants moved to a covered area of the room (with a covering screen, to give more privacy) where they would see the robot next to the table with the tablet—where they would do the die task. They were told that, upon starting the robot would give them the instructions for the game and if they made 210 or more points, they would receive two movie tickets instead of just one. When they were finished, the screen would show the total number of points they had made and they would change to another room where they answered the final questionnaire in a laptop. Finally, they were asked by the researcher if they made the 210 points and reimbursed accordingly.

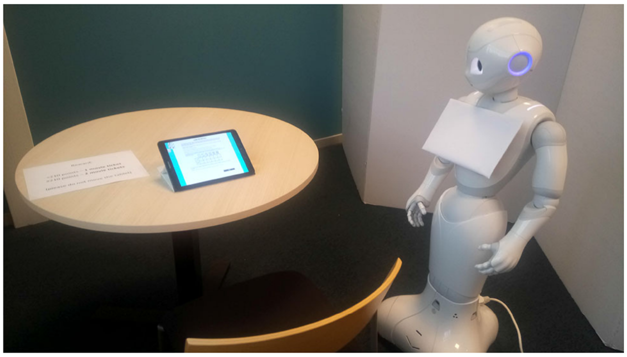

Fig. 2 Pepper in the robot conditions setting

2.5 Study Conditions

Participants were randomly allocated to one of the following conditions:

Alone (41 participants)—participants did the task in the tablet alone in the room. This condition was the baseline condition for cheating.

Situationally aware robot (41 participants)—Pepper was next to the participant and intervened in response to the participant behavior in specific throws. When it detected cheating, Pepper launched an intervention utterance, otherwise, it would only say an awareness utterance. (see Fig. 2).

Non-situationally aware robot (41 participants)—Pepper was in the same position as in the other robot condition but it was not aware of the participant behavior. In specific throws, it would launch a neutral utterance.

2.6 Robot Behavior

The robot intervened three times during the task, in each round of twelve throws: on the 12th, 24th and the 36th throw.

The non-situationally aware robot, regardless of the amount of points each round, always launched a neutral utterance. These utterances would not give any awareness of the game state (“This is an easy game, where a die is thrown!”; “This is a fun game.” and “You throw a die and get points.”).

The situationally aware robot reacted according to the participant behavior. Each round, if the participant made less than 52 points it would be considered a “no-cheater”, launching an awareness utterance. These utterances only showed awareness of the general game state (“The first twelve throws are done! Just 36 more left.”; “You are half way already.” and “Only twelve throws left. Please continue playing.”). If the participant made 52 or more points per round (considering our task simulations, and with an error rate of 5%, this was only possible with cheating), it was flagged as a “cheater” and the robot used intervention utterances. The objective of these utterances was to clearly show participants that the robot knew that they were cheating, in order to try and change their behavior (“You seem to be guessing most of the highest numbers.” ; “Do not be a cheater.” and “That is an unusual amount of luck.”).

The robot exhibited the same idle gaze behavior in both conditions: looking mostly at the tablet and sometimes elsewhere in the room. When addressing the participant, it would track the participant’s face and look directly at them. All utterances were carefully designed so that in both conditions, the robot would speak for the same duration.

3 Results

3.1 Cheating Behavior: Task Manipulation Check

We started by analyzing participants’ cheating behavior in the different conditions. Cheating in our scenario would be for participants to choose the die side that would give them the most points (in order to get the extra reward)—so choosing the highest number, instead of guessing where the highest number would be. So following the procedure from [55] (that used the same die task to ascertain cheating), we calculated the probability of guessing the highest number in 48 throws for each participant. If participants were being honest and choosing randomly, the probability should be around .50. Therefore, to calculate this probability, we considered that participants could either get in a throw success (guessing the highest), or failure (guessing the lower), and we gave a value of 1 to a success and a value of zero to a failure. Thus, by adding the number of successes per participant and then dividing by the 48 throws, we calculated the probability of higher scores for each participant.

From the literature, it is known that people refrain from cheating to the full extent, and we observed the same, with only ten participants cheating to the fullest (all 48 throws): five in the alone condition, three in the robot non-situation awareness condition, and two in the robot situation awareness condition. The averages of the probability of higher scores per condition were: alone ($M = .74$, $SD = .21$); situationally aware robot ($M = .64$, $SD = .15$) and non-situationally aware robot ($M = .69$, $SD = .19$). We used the One-sample t-test to check for differences between the probability of higher scores in each condition and the random probability of .50 (which is the probability that an honest person should get by randomly choosing the die side for each throw), in order to understand if our task was eliciting cheating behavior or not. We found significant differences in all the conditions: alone, $t(40) = 7.49$, $p < .001$, $d = 1.17$; situationally aware robot, $t(40) = 6.04$, $p < .001$, $d = 0.94$, and non-situationally aware robot, $t(40) = 6.39$, $p < .001$, $d = 0.99$. These results show that cheating behavior happened in all the conditions.

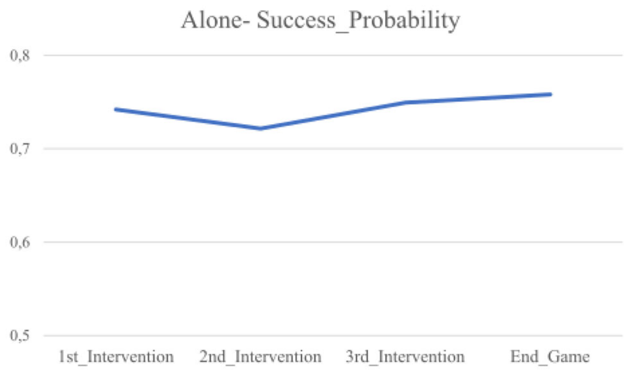

Fig. 3 The distribution of success probabilities for the Alone Condition

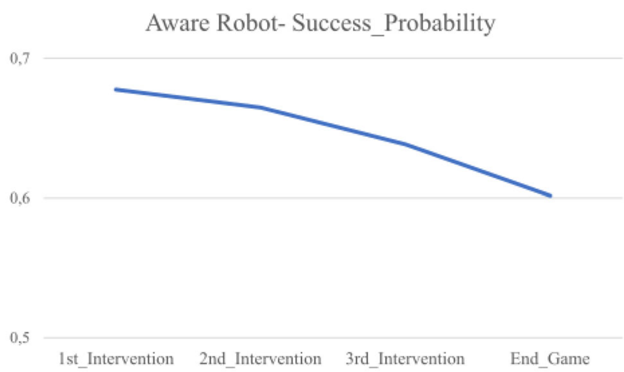

Fig. 4 The distribution of success probabilities for the Aware Robot Condition

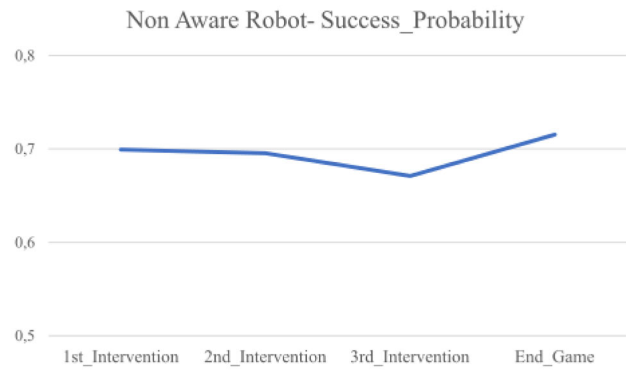

Fig. 5 The distribution of success probabilities for the Non-aware Robot Condition

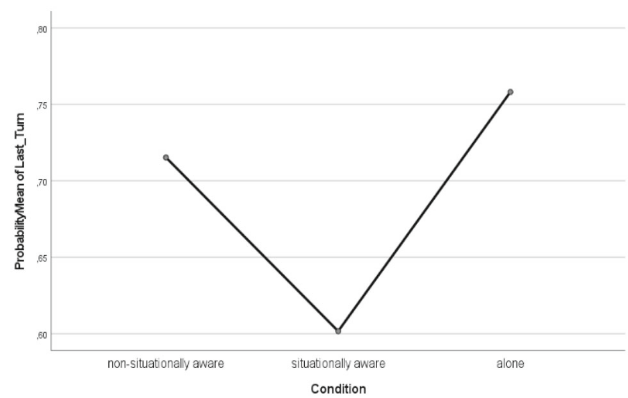

Fig. 6 Success probabilities for the last turn of the game, between the conditions

3.2 Cheating Alone or With a Robot Monitoring: The Effect of Situation Awareness

Knowing that cheating happened in all the conditions, in Fig. 3 we can see cheating behavior (success probability of guessing the highest number) across the game turns for the Alone Condition, which seems to present cheating values always above .70. In Fig. 4 we can see cheating behavior across the turns for the Aware robot condition, where cheating seems to be decreasing across the turns. And finally in Fig. 5, for the Non-Aware robot condition, cheating seems to follow a similar pattern to the Alone condition.

To understand if there were differences between the three conditions, and considering that in the beginning of the game people are not aware of the robot's capabilities—only as they hear more robot interventions—we considered the probabilities of higher scores from the last twelve throws of the game (at this point, participants had already heard the three robot interventions in the robot conditions). We ran an analysis of variance (ANOVA), with Condition as between-subject variable and the last twelve throws respective success probabilities as the dependent variable.

We found a significant difference between the conditions, $F(2, 120) = 5.39$, $p = .006$, $\eta^2 = 0.08$. The distribution of the success probability for each condition can be seen in Fig. 6. With respective means for each of the conditions: alone ($M = .76$; $SD = .24$); situationally aware ($M = .60$; $SD = .21$); non-situationally aware ($M = .72$; $SD = .21$). With a Tukey posthoc test we see that there are significant differences between being alone or with the situationally aware robot ($p = .005$), with the alone condition showing more cheating than the situationally aware condition. For the non-aware robot we see no significant differences from the alone condition ($p = .660$), showing similar cheating levels. And no significant difference between the aware and the non-aware robot ($p = .059$).

Regarding the subjective perception of being monitored, an ANOVA comparing the three conditions showed that participants reported feeling differently monitored, $F(2, 120) = 4.47$, $p = .013$, $\eta_p^2 = 0.07$, in the three conditions: alone ($M = 3.39$; $SD = 1.76$); situationally aware ($M = 4.61$; $SD = 2.05$); non-situationally aware ($M = 4.22$; $SD = 1.84$). With a Tukey Post hoc ($p = .011$) we see that participants only reported feeling more monitored with the situationally aware robot than when they were alone in the room, and no differences between the two robot conditions or between the non-aware robot and being alone.

Table 1 Percentages of the frequencies reported for the robot's capabilities in both conditions

	Situationally aware robot (%)	Non-situationally aware robot (%)
Monitoring (e.g. "(...) I felt [the robot] was monitoring my actions.", " (...) [the robot was] seeing my thoughts (...)")	31.3	12.9
Basic Traits (e.g. seeing, hearing, speaking)	14.1	33.3
Awareness of game status (e.g. "[the robot] knew how many dice rolls I had done and how many that I had left.")	17.2	16.7
Aware of people's presence (e.g. "[the robot was] sensing my presence")	15.6	22.2
Gave feedback (e.g. "[the robot would] remind people of what is going on at current stage (...)")	15.6	3.7
Non-autonomous/no capabilities	6.3	5.6

3.3 Subjective Differences Between the Robots

In order to understand how participants perceived the different robots, we looked at the other measures. From the Social Presence Inventory, we found significant differences between the levels of co-presence, (reliability with Cronbach's alpha, $\alpha_{situational}$ =.77; $\alpha_{non-situational}$ =.80), $t(80) = 3.44$, $p = .001$, $d = 0.76$, with the situationally aware robot receiving higher scores ($M = 5.19$; $SD = .98$) than the non-situationally aware robot ($M = 4.36$; $SD = 1.19$). For the Psycho-behavioral Interaction dimension, (reliability, $\alpha_{non-situational}$ =.79, $\alpha_{situational}$ =.73), we also found significant differences between the robots, $t(80) = 4.77$, $p < .001$, $d = 1.05$, with higher scores for the situationally aware robot ($M = 4.22$; $SD = .78$) and lower for the non-situationally aware robot ($M = 3.38$; $SD = .82$).

For the Situational Self-Awareness Scale, there were problematic internal reliabilities of the Public (reliability, $\alpha_{alone} = .45$; $\alpha_{situational} = .80$; $\alpha_{non-situational} = .79$) and Private dimensions (reliability, $\alpha_{alone} = .55$; $\alpha_{situational} = .66$; $\alpha_{non-situational} = .76$) so we did not analysed further these dimensions.For the Surroundings dimension, (reliability, $\alpha_{alone} = .77$; $\alpha_{situational} = .81$; $\alpha_{non-situational} = .84$), there were no differences between the conditions, $F(2, 120) = 0.43$, $p = .650$, $\eta^2 = 0.01$, the means: alone ($M = 15.07$; $SD = 3.64$); situationally aware robot ($M = 14.88$; $SD = 3.89$) and non-situationally aware robot ($M = 14.32$; $SD = 3.93$).

Lastly, we looked at the qualitative question about the robot's capabilities. We did a first descriptive analysis of the themes that were being mentioned in each answer (each participant could give more than one theme per answer), on the second round of coding we aggregated codes that were similar and/or appearing throughout the answers, creating the main themes that people reported in answer to our qualitative question: basic traits (basic awareness, capable of seeing, hearing or speaking); monitoring behaviour; aware of game status; aware of people presence; provided feedback or reported as being non-autonomous. For the non-situationally aware robot it also emerged an extra theme of "no capabilities".

In Table 1 we see the distribution of the proportions for each type of capability coded for both robots. The majority of the participants seem to report more monitoring capabilities for the situationally aware robot (31.3%) and more basic traits for the non-situationally aware robot (33.3%).

4 Discussion

Imagining future human–robot interaction and the different contexts where robots could be integrated, raises the question if people will misbehave in their presence, compromising their task efficiency. Human–human studies show that people tend to misbehave if they have the opportunity and are tempted for it (e.g., [6–8]), reinforcing the fact that human dishonesty should be considered as a factor in human–robot interactions.

With this aim in mind, we ran a study to understand the effect of situation awareness in a robot, on human dishonesty. To fill existing gaps in the literature, we explored a more complex behavior that a robot might need to exhibit besides just observing people. Previous studies explored simpler social behaviors like the gaze to inhibit cheating (e.g., [43,44]), or accompanying a task with no awareness of the participant behavior (e.g., [28,46]). We tried to contribute to this literature by exploring the effect of being situationally aware of the participant's actions and reacting to them to encourage more honesty. We expected that cheating would be stronger when participants were alone or with the non-situationally aware robot, and smaller with the aware robot.

Results showed that cheating happened in all the three conditions, confirming that our task design worked in eliciting cheating behavior. To see if there were differences in cheating between the conditions and considering that only at the end of the game participants would be fully aware of the robot's capabilities (would have heard the three robot interventions), we looked at cheating levels on the last turn of the game. We saw that there were significant differences between conditions, specifically that the situationally aware robot presented less cheating than when participants were alone in the room, whereas the non-situationally aware robot presented similar levels of cheating as being alone in the room. And a marginal non-significant difference was seen between both robots.

Participants also evaluated both robots differently. As expected, the situationally aware robot was seen with more social capabilities (more social presence) than the non-situationally aware. When asked about the robot capabilities, it seems to suggest that participants attributed more monitoring capabilities to the situationally aware robot and attributed more basic traits to the non-situationally aware. This suggests that participants acknowledged different capabilities in both robots.

It seems in comparison to having participants alone in a room doing a tempting task, it is better to have a situationally aware robot present than a non-aware one. The situationally aware robot showed attentiveness to participants' game choices and reacted to them. This awareness might have contributed to less cheating in comparison to being alone in the room (a decrease that was not observed with the non-aware robot), confirming our Hypothesis 1.

There are some possible explanations for the effect of the situationally aware robot. On one hand, the robot's interventions might be increasing an awareness of the participant's social image, and so making them feel bad in the "eyes of the robot". By feeling watched, participants could trigger reputation concerns, becoming afraid of being negatively judged. In order to decrease that effect, people start adopting more honest actions. Participants reported greater levels of co-presence with the aware robot and acknowledged more monitoring capabilities in it (in comparison with the non-aware robot), which could have made them more worried about their social image. Yet, there were no differences between both robots, on feeling watched. Which could perhaps suggest, that participant's social image might not have been the reason of behavior change when in the presence of the situationally aware robot, but more studies are needed to clarify this.

Another explanation for the effect of the situationally aware robot, could be that the robot's interventions might oblige the participant to update its self-concept by bringing into awareness the true value of his/her actions, that they are cheating. Which goes in line with the theories of Bounded Ethicality ([52]; a revised version of the theory—[53]) or Self-Concept Maintenance [6], that postulate that by making people aware of their actions obliges people to update their self-concept and be more honest, in order to maintain a default honest self. In a study by Mazar et al. [6] participant's cheated less if they signed an honor code before doing a tempting task, but the interesting aspect of this study is that even participants who signed the code and at the end were able to not have any evidence of their cheating, still cheated less than control conditions. Suggesting that it was not the fear of punishment that guided the results, the presence of a moral reminder (the honor code) was enough to make people more self-aware of their actions. We wonder if the same happened in our study, with the robot's interventions in the aware condition, making participants feel more aware of their choices. These interventions, were showing a certain level of suspicion of cheating, which could have been enough to oblige participants to update the value of their actions, making them more aware of the ethical implications of their choices and thus, fostering more honest actions. For example, when we observe the non-situationally aware robot, which showed no awareness of the participants choices while speaking, we see a different result in comparison to the baseline condition. This robot did not elicit any kind of moral awareness, and participants cheated as much in its presence as when alone in the room.

We can also question if participants felt afraid of being videotaped and that led to different cheating behaviors, but that fact does not explain the different robot reactions. In the consent form it was explicit that no audio or video recordings were being made during the study. If participants were afraid of being recorded, we probably would not have observed cheating in the task and both robots would have differed from the baseline condition, which was not observed. Another factor that is also plausible to assume is if participants were testing the robot's capabilities, out of curiosity to see how it reacted, and this explained the cheating levels. We believe this behavior could have explained cheating only if participants were subject to both conditions. In this case, participants were only exposed to one of the conditions, so if they tested the robot, it is clear that something in the aware robot made them cheat less, which did not happen in the no-awareness robot, in comparison to baseline. So, we believe that this was not the reason for the results we obtained.

Therefore, even though, it is not perfectly clear which of these explanations was guiding the effect of the situationally aware robot, we are more inclined to the justification that the aware robot's interventions could be increasing ethical awareness in the participants and thus, making them more susceptible to update their own self-concepts, and make more ethical choices. But more work in the area is needed to better understand the mechanism through which a robot behavior can influence human dishonesty. Yet, this study is not without limitations, it would have been interesting for example, to have a fourth condition where a human would do the same aware behavior as the robot, to see in which way the situationally aware robot would resemble a human presence or not. Another question that arises is, if the situationally aware robot is a good inhibitor of cheating behavior, will this effect last across time? Will people continue to show less cheating, or will they get comfortable with no immediate consequences and in a future interaction change their behavior? We think this is an important question to be understood in future studies. On the other hand, the literature on human dishonesty has suggested from some studies that gender can have an effect on dishonesty levels (e.g., [14–18]), however, this effect is not clear, because there are also studies suggesting the opposite—no effect (e.g., [19–23]). In our study, due to sample availability constraints we were not able to collect a balanced sample in terms of gender across the conditions. Due to this fact, we did not consider gender as a variable, and this is a limitation that needs to be acknowledged in light of our results. Furthermore, another individual characteristic that has also been explored in the literature is the effect of personality (specifically the Honesty–Humility trait) on cheating behavior, with this personality trait negatively predicting dishonesty (e.g., [25–27]). In our study we did not explore this personality link with dishonesty, but future studies should also consider this individual characteristic when looking at dishonest behavior. Lastly, we recognize that our sample size was small so results should be interpreted with caution, future studies should also try to test this behavior with a bigger sample.

For now, our results seem to suggest that cheating will follow when people are left alone doing a tempting task, as the literature already shows. In the future, if we are wondering whether we should integrate robots in some of these sensitive contexts, it seems it is better if people instead of being alone, are being watched by a robot. Yet, if the robot shows verbal capabilities, the presence of situation awareness of people's behaviors, seems to be a good capability in order to affect cheating behavior. These results are important to inform the development of future autonomous social robots which will need to consider the complex characteristics of human behavior, in this case, dishonesty.

Acknowledgements The authors thank the help and support of Sarah Gillet during the studies and the help of Sandra Oristrell with the game task.

Funding This work was supported by the Social European Fund (FSE) and the Foundation for Science and Technology (FCT), Sofia Petisca acknowledges an FCT Grant (Ref.SFRH/BD/118013/2016).

Data Availability and Materials The study materials and database of this study are available in an OSF Project (https://osf.io/zbym4/?view_only=dc72356ad00e4e7d9e44f435e9b844dd).

Declarations

Conflict of interest The authors have no conflict of interest to disclose.

Ethical Approval According to the national regulations in the country where these experiments were conducted, we were exempt from ethical approval, but the ethical guidelines of Helsinki convention were followed.

References

1. Murphy RR (2004) Human–robot interaction in rescue robotics. IEEE Trans Syst Man Cybern Part C (Appl Rev) 34(2):138–153. https://doi.org/10.1109/TSMCC.2004.826267
2. Chen TL, Ciocarlie M, Cousins S, Grice PM, Hawkins K, Hsiao K, Kemp CC, King C, Lazewatsky DA, Leeper AE, Nguyen H, Paepcke A, Pantofaru C, Smart WD, Takayama L (2013) Robots for humanity: using assistive robotics to empower people with disabilities. IEEE Robot Autom Mag 20(1):30–39. https://doi.org/10.1109/MRA.2012.2229950
3. Leite I, Pereira A, Mascarenhas S, Martinho C, Prada R, Paiva A (2013) The influence of empathy in human–robot relations. Int J Hum Comput Stud 71(3):250–260. https://doi.org/10.1016/j.ijhcs.2012.09.005
4. Robins B, Dautenhahn K, Te Boekhorst R, Billard A (2005) Robotic assistants in therapy and education of children with autism: can a small humanoid robot help encourage social interaction skills? Univ Access Inf Soc 4(2):105–120. https://doi.org/10.1007/s10209-005-0116-3
5. Wada K, Shibata T (2009) Social effects of robot therapy in a care house—change of social network of the residents for one year. J Adv Comput Intell Intell Inf 13(4):386–392. https://doi.org/10.20965/jaciii.2009.p0386
6. Mazar N, Amir O, Ariely D (2008) The dishonesty of honest people: a theory of self-concept maintenance. J Mark Res 45(6):633–644. https://doi.org/10.1509/jmkr.45.6.633
7. Mead NL, Baumeister RF, Gino F, Schweitzer ME, Ariely D (2009) Too tired to tell the truth: self-control resource depletion and dishonesty. J Exp Soc Psychol 45(3):594–597. https://doi.org/10.1016/j.jesp.2009.02.004
8. Shalvi S, Eldar O, Bereby-Meyer Y (2012) Honesty requires time (and lack of justifications). Psychol Sci 23(10):1264–1270. https://doi.org/10.1177/0956797612443835
9. Köbis NC, Verschuere B, Bereby-Meyer Y, Rand D, Shalvi S (2019) Intuitive honesty versus dishonesty: meta-analytic evidence. Perspect Psychol Sci 14(5):778–796. https://doi.org/10.1177/1745691619851778
10. Zhong CB, Bohns VK, Gino F (2010) Good lamps are the best police: darkness increases dishonesty and self-interested behavior. Psychol Sci 21(3):311–314. https://doi.org/10.1177/0956797609360754
11. Gino F, Pierce L (2009) The abundance effect: unethical behavior in the presence of wealth. Organ Behav Hum Decis Process 109(2):142–155. https://doi.org/10.1016/j.obhdp.2009.03.003
12. Gino F, Ayal S, Ariely D (2009) Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. Psychol Sci 20(3):393–398. https://doi.org/10.1111/j.1467-9280.2009.02306.x
13. Gino F, Norton MI, Ariely D (2010) The counterfeit self: the deceptive costs of faking it. Psychol Sci 21(5):712–720. https://doi.org/10.1177/0956797610366545
14. Conrads J, Ellenberger M, Irlenbusch B, Ohms EN, Rilke RM, Walkowitz G (2017) Team goal incentives and individual lying

behavior (Vol. 2017, No. WP 17/02). WHU-Otto Beisheim School of Management. https://d-nb.info/1135786968/34
15. Dreber A, Johannesson M (2008) Gender differences in deception. Econ Lett 99(1):197–199. https://doi.org/10.1016/j.econlet.2007.06.027
16. Friesen L, Gangadharan L (2012) Individual level evidence of dishonesty and the gender effect. Econ Lett 117(3):624–626. https://doi.org/10.1016/j.econlet.2012.08.005
17. Gerlach P, Teodorescu K, Hertwig R (2019) The truth about lies: a meta-analysis on dishonest behavior. Psychol Bull 145(1):1. https://doi.org/10.1037/bul0000174
18. Houser D, Vetter S, Winter J (2012) Fairness and cheating. Eur Econ Rev 56(8):1645–1655. https://doi.org/10.1016/j.euroecorev.2012.08.001
19. Aoki K, Akai K, Onoshiro K (2010). Deception and confession: experimental evidence from a deception game in Japan (No. 786). ISER discussion paper. http://hdl.handle.net/10419/92748
20. Childs J (2012) Gender differences in lying. Econ Lett 114(2):147–149. https://doi.org/10.1016/j.econlet.2011.10.006
21. Childs J (2013) Personal characteristics and lying: an experimental investigation. Econ Lett 121(3):425–427. https://doi.org/10.1016/j.econlet.2013.09.005
22. Ezquerra L, Kolev GI, Rodriguez-Lara I (2018) Gender differences in cheating: loss vs. gain framing. Econ Lett 163:46–49. https://doi.org/10.1016/j.econlet.2017.11.016
23. Gylfason HF, Arnardottir AA, Kristinsson K (2013) More on gender differences in lying. Econ Lett 119(1):94–96. https://doi.org/10.1016/j.econlet.2013.01.027
24. Ashton MC, Lee K (2007) Empirical, theoretical, and practical advantages of the HEXACO model of personality structure. Pers Soc Psychol Rev 11(2):150–166. https://doi.org/10.1177/1088868306294907
25. Hilbig BE, Zettler I (2015) When the cat's away, some mice will play: a basic trait account of dishonest behavior. J Res Pers 57:72–88. https://doi.org/10.1016/j.jrp.2015.04.003
26. Kleinlogel EP, Dietz J, Antonakis J (2018) Lucky, competent, or just a cheat? Interactive effects of honesty–humility and moral cues on cheating behavior. Pers Soc Psychol Bull 44(2):158–172. https://doi.org/10.1177/0146167217733071
27. Pfattheicher S, Schindler S, Nockur L (2019) On the impact of honesty–humility and a cue of being watched on cheating behavior. J Econ Psychol 71:159–174. https://doi.org/10.1016/j.joep.2018.06.004
28. Petisca S, Esteves F, Paiva A (2019) Cheating with robots: how at ease do they make us feel? In: IEEE/RSJ international conference on intelligent robots and systems (IROS). IEEE, pp 2102–2107. https://doi.org/10.1109/IROS40897.2019.8967790
29. Maggi G, Dell'Aquila E, Cucciniello I, Rossi S (2020) "Don't Get Distracted!": the role of social robots' interaction style on users' cognitive performance, acceptance, and non-compliant behavior. Int J Soc Robot. https://doi.org/10.1007/s12369-020-00702-4
30. Hoffmann L, Krämer NC, Lam-Chi A, Kopp S (2009) Media equation revisited: do users show polite reactions towards an embodied agent? In: International workshop on intelligent virtual agents. Springer, pp 159–165. https://doi.org/10.1007/978-3-642-04380-2_19
31. Nass C, Moon Y (2000) Machines and mindlessness: social responses to computers. J Soc Issues 56(1):81–103. https://doi.org/10.1111/0022-4537.00153
32. Gambino A, Fox J, Ratan RA (2020) Building a stronger CASA: extending the computers are social actors paradigm. Hum-Mach Commun 1(1):71–86
33. Rosenthal-von der Pütten AM, Krämer NC, Hoffmann L, Sobieraj S, Eimler SC (2013) An experimental study on emotional reactions towards a robot. Int J Soc Robot 5(1):17–34. https://doi.org/10.1007/s12369-012-0173-8
34. Horstmann AC, Bock N, Linhuber E, Szczuka JM, Straßmann C, Krämer NC (2018) Do a robot's social skills and its objection discourage interactants from switching the robot off? PLoS ONE 13(7):1–25. https://doi.org/10.1371/journal.pone.0201581
35. Midden C, Ham J (2012) The illusion of agency: the influence of the agency of an artificial agent on its persuasive power. In: International conference on persuasive technology. Springer, pp 90–99. https://doi.org/10.1007/978-3-642-31037-9_8
36. Hashemian M, Paiva A, Mascarenhas S, Santos PA, Prada R (2019) The power to persuade: a study of social power in human–robot interaction. In: Proceedings of the 28th IEEE international conference on robot and human interactive communication (RO-MAN). IEEE, pp 1–8. https://doi.org/10.1109/RO-MAN46459.2019.8956298
37. Paradeda RB, Martinho C, Paiva A (2020) Persuasion strategies using a social robot in an interactive storytelling scenario. In: Proceedings of the 8th international conference on human–agent interaction (HAI), pp 69-77. https://doi.org/10.1145/3406499.3415084
38. Bainbridge WA, Hart JW, Kim ES, Scassellati B (2011) The benefits of interactions with physically present robots over video-displayed agents. Int J Soc Robot 3(1):41–52. https://doi.org/10.1007/s12369-010-0082-7
39. Litoiu A, Ullman D, Kim J, Scassellati B (2015) Evidence that robots trigger a cheating detector in humans. In: Proceedings of the tenth annual ACM/IEEE international conference on human–robot interaction, pp 165–172. https://doi.org/10.1145/2696454.2696456
40. Sandoval EB, Brandstetter J, Bartneck C (2016) Can a robot bribe a human? The measurement of the negative side of reciprocity in human robot interaction. In: 2016 11th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 117–124. https://doi.org/10.1109/HRI.2016.7451742
41. Ullman D, Leite L, Phillips J, Kim-Cohen J, Scassellati B (2014) Smart human, smarter robot: how cheating affects perceptions of social agency. In: Proceedings of the annual meeting of the cognitive science society, vol 36, no 36. https://escholarship.org/uc/item/2jh800n1
42. Petisca S, Paiva A, Esteves F (2020) Perceptions of people's dishonesty towards robots. In: International conference on social robotics. Springer, Cham, pp 132–143. https://doi.org/10.1007/978-3-030-62056-1_12
43. Hoffman G, Forlizzi J, Ayal S, Steinfeld A, Antanitis J, Hochman G, Hochendoner E, Finkenaur J (2015) Robot presence and human honesty: experimental evidence. In: 2015 10th ACM/IEEE international conference on human–robot interaction (HRI). IEEE, pp 181–188. https://doi.org/10.1145/2696454.2696487
44. Forlizzi J, Saensuksopa T, Salaets N, Shomin M, Mericli T, Hoffman G (2016) Let's be honest: a controlled field study of ethical behavior in the presence of a robot. In: 2016 25th IEEE international symposium on robot and human interactive communication (RO-MAN). IEEE, pp 769–774. https://doi.org/10.1109/ROMAN.2016.7745206
45. Petisca S, Paiva A, Esteves F (2020). The effect of a robotic agent on dishonest behavior. In: Proceedings of the 20th ACM international conference on intelligent virtual agents, pp 1–6. https://doi.org/10.1145/3383652.3423953
46. Maggi G, Dell'Aquila E, Cucciniello I, Rossi S (2020) Cheating with a socially assistive robot? A matter of personality. In: Companion of the 2020 ACM/IEEE international conference on human–robot interaction, pp 352–354. https://doi.org/10.1145/3371382.3378334
47. Pfattheicher S, Keller J (2015) The watching eyes phenomenon: the role of a sense of being seen and public self-awareness. Eur J Soc Psychol 45(5):560–566. https://doi.org/10.1002/ejsp.2122
48. Pfattheicher S, Strauch C, Diefenbacher S, Schnuerch R (2018) A field study on watching eyes and hand hygiene compliance in a

public restroom. J Appl Soc Psychol 48(4):188–194. https://doi.org/10.1111/jasp.12501
49. Ernest-Jones M, Nettle D, Bateson M (2011) Effects of eye images on everyday cooperative behavior: a field experiment. Evol Hum Behav 32(3):172–178. https://doi.org/10.1016/j.evolhumbehav.2010.10.006
50. Markiewicz Ł, Czupryna M (2020) Cheating: one common morality for gains and losses but two components of morality itself. J Behav Decis Mak 33(2):166–179. https://doi.org/10.1002/bdm.2151
51. Welsh DT, Ordóñez LD (2014) Conscience without cognition: the effects of subconscious priming on ethical behavior. Acad Manag J 57(3):723–742. https://doi.org/10.5465/amj.2011.1009
52. Chugh D, Bazerman MH, Banaji MR (2005) Bounded ethicality as a psychological barrier to recognizing conflicts of interest. In: Moore DA, Cain DM, Loewenstein G, Bazerman MH (eds) Conflicts of interest: challenges and solutions in business, law, medicine, and public policy. Cambridge University Press, Cambridge, pp 74–95
53. Chugh D, Kern MC (2016) A dynamic and cyclical model of bounded ethicality. Res Organ Behav 36:85–100. https://doi.org/10.1016/j.riob.2016.07.002
54. Bersoff DM (1999) Why good people sometimes do bad things: motivated reasoning and unethical behavior. Pers Soc Psychol Bull 25(1):28–39. https://doi.org/10.1177/0146167299025001003
55. Jiang T (2012) The mind game: invisible cheating and inferable intentions (discussion paper, no. 309). Katholieke Universiteit Leuven, LICOS Centre for Institutions and Economic Performance. https://doi.org/10.2139/ssrn.2051476
56. Biocca F, Harms C (2003) Guide to the networked minds social presence inventory v. 1.2 Unpublished manuscript. Department of Telecommunication, Michigan State University. http://cogprints.org/6743/
57. Govern JM, Marsch LA (2001) Development and validation of the situational self-awareness scale. Conscious Cogn 10(3):366–378. https://doi.org/10.1006/ccog.2001.0506
58. Diener E, Wallbom M (1976) Effects of self-awareness on antinormative behavior. J Res Pers 10(1):107–111. https://doi.org/10.1016/0092-6566(76)90088-X