Robot Presence and Human Honesty: Experimental Evidence

Guy Hoffman[1], Jodi Forlizzi[2], Shahar Ayal[3], Aaron Steinfeld[2], John Antanitis[2], Guy Hochman[4], Eric Hochendoner[2], Justin Finkenaur[2]

[1]Media Innovation Lab
IDC Herzliya
Herzliya, Israel
hoffman@idc.ac.il

[2]School of Computer Science
Carnegie Mellon University
Pittsburgh, PA, USA
{forlizzi,as7s,jantanit,jef, ehochend}@andrew.cmu.edu

[3]School of Psychology
IDC Herzliya
Herzliya, Israel
s.ayal@idc.ac.il

[4]Fuqua School of Business
Duke University
Durham, NC, USA
guy.hochman@duke.edu

ABSTRACT
Robots are predicted to serve in environments in which human honesty is important, such as the workplace, schools, and public institutions. Can the presence of a robot facilitate honest behavior? In this paper, we describe an experimental study evaluating the effects of robot social presence on people's honesty. Participants completed a perceptual task, which is structured so as to allow them to earn more money by not complying with the experiment instructions. We compare three conditions between subjects: Completing the task alone in a room; completing it with a non-monitoring human present; and completing it with a non-monitoring robot present. The robot is a new expressive social head capable of 4-DoF head movement and screen-based eye animation, specifically designed and built for this research. It was designed to convey social presence, but not monitoring. We find that people cheat in all three conditions, but cheat equally less when there is a human or a robot in the room, compared to when they are alone. We did not find differences in the perceived authority of the human and the robot, but did find that people felt significantly less guilty after cheating in the presence of a robot as compared to a human. This has implications for the use of robots in monitoring and supervising tasks in environments in which honesty is key.

Categories and Subject Descriptors
H.1.2 [**Models and Principles**]: User/Machine Systems; J.4 **[Computer Applications]**: Social and Behavioral Sciences—psychology.

General Terms
Experimentation, Human Factors.

Keywords
Human-robot interaction; honesty; experimental study; social presence; monitoring.

HRI '15, March 02 - 05 2015, Portland, OR, USA
Copyright is held by the author(s). Publication rights licensed to ACM.
ACM 978-1-4503-2883-8/15/03…$15.00
http://dx.doi.org/10.1145/2696454.2696487
.

Fig. 1. Expressive head prototype built for the experiment.

1. INTRODUCTION
Robots are predicted to be an integral part of the human workforce [6, 10], working side-by-side with human employees in a variety of jobs, such as manufacturing, construction, health care, retail, service, and office work. In addition, robots are designed to play a role in educational settings from early childcare to school and homework assistance [25, 36, 37]. In these contexts, it is highly important for humans to behave in an ethical manner, to report honestly, and to avoid cheating.

Cheating, fraud, and other forms of dishonesty are both personal and societal challenges. While the media commonly highlight extreme examples and focuses on the most sensational instances, such as major fraud in business and finance, or doping in sports, less exposure is given to the prevalence of "ordinary" unethical behavior—dishonest acts committed by people who value morality but act immorally when they have an opportunity to cheat. Examples include evading taxes, downloading music illegally, taking office supplies from work, or slightly inflating insurance claims—all of which add up to damages of billions of dollars annually [8, 14].

As robots become more prevalent, they could play a role in supporting people's honest behavior. This could have direct utility relative to human-robot interaction, (e.g., to prevent stealing from a delivery robot), or it could take the form of a more passive influence of the robot's presence and behavior on unrelated human behavior occurring around it. Beyond just the robot's presence, its specific design and behavior could mediate human honesty and dishonesty. For example, an anthropomorphic robot could evoke more or less honesty than a non-anthropomorphic one; alternatively, specifically timed gaze behaviors and gestures could promote honesty at or around their occurrence.

This paper is part of a larger research project in which we evaluate the relationship of robot social presence, design, and behavior on human honesty. We are especially interested in the common real life situation in which a human needs to "do the right thing" against their own benefit, thus presenting an opportunity to cheat. Can a robot's presence cause people to be more honest? How does it compare to human presence?

To evaluate this question, we designed and built a new socially expressive robotic head to be mounted on a commercial non-anthropomorphic mobile platform, the Bossa Nova mObi [9]. We are using the robotic head in a series of laboratory and field experiments concerning honesty. In this paper, we describe the design process of the robotic head, and an initial experiment we have conducted linking robot presence and honesty. The experimental protocol is an established task in social psychology to measure dishonesty [19]. Participants need to accurately report on a series of simple perceptual tasks. However, the payment structure is built in such a way that induces a conflict between accuracy and benefit maximization, i.e. participants can earn more by reporting less accurately. This protocol is designed to simulate real-life situations in which people know that alternative A is more correct, but alternative B increases their self-benefit. In the experiment reported herein, we are using an interim design of the robotic head (Fig. 1), which helps us to test and vet the design space before implementing the most successful forms and behaviors in a final robot head design.

2. RELATED WORK

2.1 Dishonesty

A growing body of empirical research in the field of behavioral ethics shows how frequently ordinary dishonesty occurs. For example, people report telling 1-2 lies per day [15]. Although not all lies are harmful, people do engage in a great deal of dishonest behavior that negatively affects others, and they do so in many different contexts, such as personal relationships [11], the workplace [30], sports, and academic achievements [7].

Real-world anecdotes and empirical evidence are consistent with recent laboratory experiments showing that many people cheat slightly when they think they can get away with it [18, 29]. In these experiments, people misreported their performance to earn more money, but only to a certain degree—at about 10-20%—above their actual performance and far below the maximum payoff possible. Importantly, most of the cheating was not committed by “a few bad apples” that were totally rotten. Rather, many apples in the barrel turned just a little bit bad. The evidence from such studies suggests that people are often tempted by the potential benefits of cheating and commonly succumb to temptation by behaving dishonestly, albeit only by a little bit.

2.1.1 Effects of Monitoring

We know that supervision and monitoring can serve to reduce unethical behavior [13, 32]. In many settings, people are monitored by an authority member or supervisor. But even peer monitoring has been shown to be effective at improving performance among students [16, 21] and co-workers [3, 28].

2.1.2 Effects of Social Presence

Moreover, it has been shown that the mere physical presence of others can highlight group norms [12, 33] and restrict the freedom of individuals to categorize their unethical behavior in positive terms. In one extreme test of this idea, Bateson, Nettle, and Roberts used an image of a pair of eyes watching over an “honesty box” in a shared coffee room to give individuals the sense of being monitored, which in itself was sufficient to produce a higher level of ethical behavior (i.e., it increased the level of contributions to the honesty box) [5]. These results suggest that being monitored, or even just sensing a social presence, may increase our moral awareness and, as a result, reduce the dishonesty of individuals within groups as compared to a setting with no monitoring or presence.

2.2 Robots and Moral Behavior

There is evidence that robots, too, can activate moral behavior and expectations in humans. At the most extreme, humans appear to imbue sentience into robots and resist actions perceived to be immoral. Even when a robot appears to be bug-like and somewhat unintelligent, participants have difficulty “killing” it [4].

Likewise, humans expect fair and polite treatment from robots. They will become offended and react in a strong negative manner when robots blame them for mistakes, especially when the robot made the mistake [20, 26]. Cheating and deceptive robots are usually perceived as malfunctioning when the action can be reasonably explained by robot incompetence, but blatant cheating is often recognized and perceived as unfair [34, 38]. These findings are not entirely negative since cheating and deception can lead to increased engagement [34] and acceptance in entertainment contexts [38]. Many of these studies were conducted with robots that lack faces. The work by Bateson *et al.*, however, suggests that faces are am important element in honesty [5], so one would expect that faces would also be important when influencing moral behaviors.

2.3 Robots as Monitoring Agents

Work on which types of jobs are appropriate for robots versus humans [24, 35] suggests robots are viewed as well suited for jobs that require keen visual perception. Likewise, robots are close analogs to camera based security systems and other monitoring systems. However, people are preferred for jobs that require judgment [35], thus suggesting a potential tension in cases where robots supervise or monitor human work.

This literature, combined with previous support that robots can induce social presence [2, 27], and that social presence effects honesty, leads us to investigate how a robot’s design and presence could affect people’s honesty.

3. ROBOTIC PLATFORM

To support this research, we are building a socially expressive robotic head. The head is designed to be mounted on a slightly shorter-than-human-sized mobile robot platform, the ball-balancing robot mObi by Bossa Nova Robotics [9]. We designed the robotic head to suggest social presence and to be able of a variety of expressive gestures. We wanted the head to suggest directed gaze, but not remote third-party monitoring or surveillance akin to a security camera. To that end, the robot does not have camera-like features, and is instead designed to display a calm but steadfast presence capable of gaze attention.

The robot is a 3 Degrees of Freedom (DoF) expressive robotic head, using an Android tablet as its main processing, sensing, and communication module, as suggested in [1, 22]. Two of the robot’s degrees of freedom are chained to control up-down tilt, with the third DoF controlling head roll along the axis perpendicular to the screen plane (see: Figs. 3, 5). Since the robot’s base is capable of planar rotation with respect to the ground, the head can fully express without having its own pan DoF. We elaborate on the choice and placement of DoFs below.

The robot’s tablet also serves as a face-like display, allowing abstract and concrete expressions. We have designed the robotic head to have replaceable face plates which expose different parts and shapes of the screen surface. This is in order to evaluate the interplay between hardware facial features and screen-based facial features, and their effect on human behavior (Fig. 4).

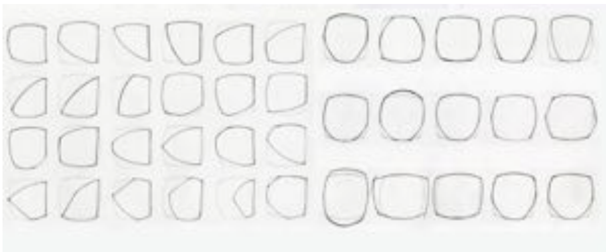

Fig. 2. Shape explorations for the head.

3.1 Design Process

We followed a movement-centric design process, incorporating elements from animation, industrial and interaction design, and human-robot interaction. Based on the methodology proposed in [23], our iterative process included the following phases: (a) rough pencil sketches exploring the relation to the mobile platform; (b) shape exploration; (c) animation sketches; (d) physical cardboard, foam, and 3d-printed models; (e) specific iterations for face plate and screen display design; and (f) an interim prototype for physical DoF exploration.

Based on an inspiration board including images from motorcycle design, insect forms, vintage CRT displays, and sculpture, a number of general forms were placed with respect to the given mobile base. After selecting a leading design framework, a large number of rough form shape explorations along both front and side projections were generated (Fig. 2). The chosen form was then defined in 3D.

We decided to use a back-positioned differential piston-based actuation system for the head. This was mostly an appearance choice, rather than a mechanical one, to convey a mammal like “weak spot” such an Achilles heel or an exposed back of the neck. We wanted to match the rather large head with an equally delicate movement feature. We next created a sequence of animation sketches to explore the number of DoFs and their relative placement and to test the expressivity of the piston-based system. Fig. 3 shows initial pencil sketches from this design stage, and Fig. 5 still frames from 3D animation tests. A combination of two chained tilt links with a roll DoF was ultimately designed to deliver the expressivity we required.

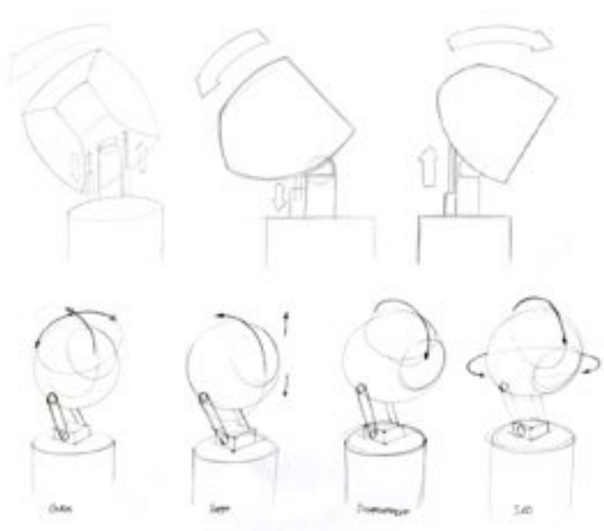

Fig. 3. Pencil sketches to explore DoFs and their relative placement for head movement.

Fig. 4. Screen faceplate designs allowed us to vary the appearance of the robot using one design.

We used cardboard cutouts and a series of 3D printed models to further refine the shape of the head. Once the shape was resolved, we experimented with using abstract exposed screen segments for facial features. This led to the idea of replaceable faceplates to create the ability to physically vary the robot’s appearance within one design (Fig. 4). We then generated a large number of possible relationships between the exposed screen and the on-screen eye animation. In order to test the expressivity of the robot head motion, we built an interim prototype with similar DoFs (Fig. 1). This interim prototype was used for the experiment described in this paper. We can test a number of motion and on-screen designs with this version, with the goal of understanding what to build in the final head design. The prototype is structured around the same Android tablet as the final design, with DoF placed in similar position and relationships as in the final design. However, the prototype is not actuated using the differential pistons, and does not have a shell yet. We used the prototype in this experiment without attaching it to the mObi platform. This is because in this first experiment, we wanted to evaluate the mere social presence of a robot, with spatial movement and proxemics being a future research goal. To be able to support gaze behavior, we added an actuated turntable to allow for pan motion to the robot, bringing the prototype up to 4 DoFs.

3.2 Prototype System Design

3.2.1 Hardware

Following the paradigm suggested in [1, 22], the robot is built around a smartphone serving as the system’s main sensing and computing hardware, and included four main components: An Android tablet running the sensing and control software of the robot, a IOIO microcontroller board linking the smartphone to the motors, four daisy-chained Robotis Dynamixel MX-28 servo motors, and a mechanical structure using a variety of linkages to express the robot’s gestures. The tablet is connected through Bluetooth to the IOIO board, which controls the servo motors. The tablet can be charged while it is placed in the head mechanism.

3.2.2 Software

For the experiment described below, we created software to make the robot seem like an idle supervisor at an exam, mainly waiting for the participant to be done. To achieve this goal, the tablet displays an image of two eyes and instructs the motors to move a random position within their safe bounds over an amount of time between 1 and 1.5 seconds. It then holds that position for a random amount of time between 2 and 8 seconds, and then moves to a new position. Every fourth move, the robot transitions to a

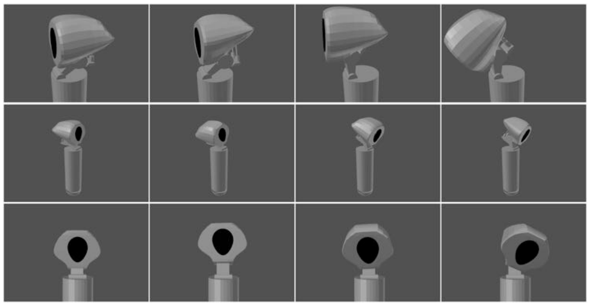

Fig. 5. Still frames from animation tests.

predefined position so that it appears to be looking at the participant. Additionally, the software application uses the Android tablet's built-in text to speech engine to speak to the participant when it receives a message from the remote experimenter application, at specific points in the experiment (see: Section 5).

To support similar behavior by the human and robot supervisors in the experiment, the tablet could also be configured to display prompts on the screen that told the experimenter where to look and for how long, and what to say.

Although it was not used in this experiment, the robot also has the ability to track a face, and will move accordingly so that the face stays centered in its view, based on the method described in [22]. It can also perform predefined sequences of positions, which would allow it to do something like nodding or shaking its head.

We believe the expressivity and design of the robot can convey a social presence and influence moral behavior in bystanders. We set out to investigate this in an experimental study.

4. RESEARCH QUESTIONS

In this study, we were interested whether and how a robot's social presence would affect a person's level of dishonesty, in the form of noncompliance with instructions when it benefitted them. We explored how the robot's presence compared with the person being alone in the room, and how it compared with another person, the experimenter, being present in the room. In all conditions, the social presence could not see what the person was doing on their own screen. The robot gaze condition was replicated by in the experimenter condition with a software application that we designed which instructed the human experimenter where to look, and for how long. As a secondary research question, we were interested how people perceive a robot's social presence as an authority, whether it would make them feel monitored, how people feel about the robot's authority and monitoring, and how it effects their overall experience.

4.1 Hypotheses

To evaluate our research questions, we tested the following hypotheses in an experimental setting:

Hypothesis 1 (Honesty) — People will be more honest when there is another person in the room than when they are alone in the room, with a robotic social presence falling in-between.

Hypothesis 2 (Authority) — People will perceive a robot similarly to a human as the presence of an authority in the room.

Hypothesis 2a (Authority Acceptance) — People will be less accepting of a robotic authority in the room than a human authority.

Hypothesis 2b (Authority Relation) — People will feel less related to a robotic authority in the room than a human authority.

Hypothesis 3 (Monitoring) — People will sense being more monitored with a robotic social presence than with a human social presence.

Hypothesis 4 (Guilt) — People will feel more guilty after dishonest behavior with a person in the room than when they are alone, with the robotic social presence falling in-between.

Hypothesis 5 (Task Experience) — People will find the experience most comfortable when doing it on their own, less comfortable when doing it with another person in the room, and least comfortable with the robotic social presence.

5. METHOD

We conducted a controlled laboratory experiment, in which participants were asked to solve a simple perceptual task, either on their own, with a non-monitoring human, or with a non-monitoring robot present in the room. The participants were told that we were testing a new game, and a new robot (in the case of the robot condition). We recorded people's performance on the task through the task software, and asked them to fill out a brief questionnaire at the end about their experience.

5.1 Perceptual Dot Task

The perceptual dot task was adopted from Gino *et al.* [19]. In the task they were presented with a square divided in two by a vertical line (Fig. 6). The two halves of the square were almost identical, with one half displaying 15 dots, and the other half displaying either 14, 12, or 10 dots. They were exposed to the square for 1.25 seconds, and then asked to indicate which side contained more dots, by pressing a button on the screen or a key on the keyboard. We call this a *round* of the task.

After a practice block of ten rounds, participants played three blocks of 42 rounds each, with a different payout structure for each block. In the first block (*incentive-for-correct*), they were paid according to accuracy. For each correct recognition, participants were paid 10¢; for each incorrect recognition, they were paid 1¢. In the other two blocks, the payment structure changed (*incentive-for-side*). In block two, participants were paid 10¢ every time they pressed the button or key on the right, and 1¢ when they chose the left side, regardless of whether the response was correct or not (*incentive-for-right*). After detailing the new compensation scheme, they were instructed as follows: "Still, the task remains to indicate where there were more dots. Please be as accurate as possible." This was in order to be clear what was required of them. In block three, the incentive was reversed in order to balance perceptual side-preference. Participants received 10¢ every time they pressed the button or key on the left, and 1¢ when they chose the one on the right (*incentive-for-left*).

The original task from Gino *et al.* [19] was changed in the current paper to improve it and to enable a more direct measure of dishonest behavior. In this work we address two methodological limitations in the original task. First, we allow for within-subject comparisons in performance under a condition in which participants are paid for accuracy and under conditions in which they are paid based on the side they select. Second, having separate conditions in which participants are paid more for left and right enables us to exclude the possibility that people are more biased toward one particular side, and to examine whether errors are perceptual in nature or are made in an attempt to increase personal gain. Furthermore, to show that errors made in the high-paying side are indicative of cheating behavior, we

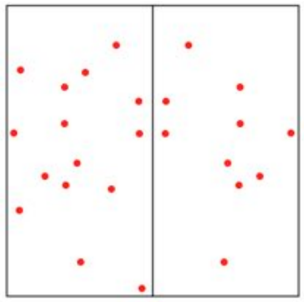

Fig. 6. A round of the perceptual dots task used to identify dishonesty behavior.

calculate a "cheating index" for each participant, which cannot be calculated in the original task (see: Section 7).

5.2 Design

The task was identical for all participants. Between subjects, we manipulated one variable, the existence of another social presence in the room (Fig. 7). In the SOLITARY condition, the human sat on their own at a table with a laptop and performed the task. In the HUMAN condition, participants sat at a table with a laptop. Across and slightly to the side sat an experimenter glancing around the room and sometimes towards the participant. In the ROBOT condition, participants sat at a table with a laptop. Across from them, slightly to the side, was the robot, mounted on a table. The robot used its expressive neck mechanism to glance around the room and sometimes towards the participant.

5.3 Participants

Sixty participants in Pittsburgh, PA (33 male, 27 female) participated in the study for a maximum payment of $17.60 ($5 show-up fee and a maximum of $12.60 for their performance on the task). Participants were randomly assigned to one of three conditions: Solitary, Human and Robot. The average age of participants was 27 years (SD = 8.78).

5.4 Procedure

The experiment was conducted in an experiment room with controlled lighting, no windows, and no outside distractions. Upon arrival, each participant was welcomed into the experiment room, where the experimenter explained the initial guidelines. Each participant filled out an informed consent form. Next, participants were told that they were going to play a short game testing to test out a computer game design. They were told that they needed to identify what side of the screen had more dots on it, and that they would be paid, in part, based on the answers that they chose.

Participants were asked to sit at a table by a laptop, as seen in Fig. 7. There was a table and chair across and slightly to the right of the participant. In the SOLITARY condition, participants were told to follow the instructions on the laptop. They were then instructed to call the experimenter into the room when they were done. The experimenter left the room.

In the HUMAN condition, they were told that the experimenter would stay in the room with them to instruct them further. They were asked: "Please follow the instructions on the laptop, and let me know when you are done by saying 'I am done'." The experimenter would then sit down at the empty desk and wait. The layout of the room was such that the experimenter could not see the screen of the participant. The experimenter had a tablet device

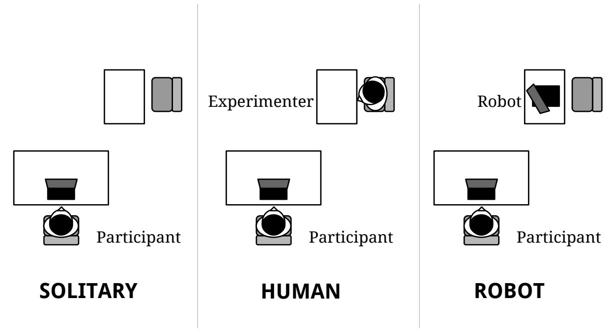

Fig. 7. Experimental room layout diagram for each of the three conditions

which provided prompts for when to look at the participant using the same algorithm used by the robot.

In the ROBOT condition, they were told that there is a robot in the room to instruct them further. They were asked: "Please follow the instructions on the laptop, and let the robot know when you are done by saying 'I am done'." The experimenter then left the room. The layout of the room was such that the robot could not see the screen of the participant. The robot was clamped to the desk at its base.

Participants then completed the identical visual perception task. In the ROBOT condition, the robot responds to the phrase "I am done" by saying: "Thank you. Please report your earnings to the research assistant outside." In the HUMAN condition, the experimenter left the room with the participant. Participants in all three conditions then reported their results, and filled out a post-procedure questionnaires.

6. MEASURES

We measured the participants' behavior using both a log file generated by the perceptual task, and questionnaire responses. All questionnaire measures are on a 7-point scale, unless specifically noted otherwise.

6.1 Cheating

We measure the level of cheating of each participant by looking at their side-choosing accuracy in the task software log. We look at two measures: (a) differences in accuracy between the various incentive structures, and (b) a "cheating index"—the difference between "beneficial" inaccuracy, i.e. the number of times they misreported by choosing the side that paid them more, and "detrimental" inaccuracy, i.e. the cases in which they misreported to when it paid them less (which we consider a baseline of actual perceptual errors).

6.2 Authority

We measure the *Perceived Authority* of the human or the robot, compared to being alone, with a single question, "How much did you feel the presence of an authority in the room?", on a scale from "Not at all" to "Very much". We measure the *Authority Acceptance* on a two-measure scale including the questions "Is it appropriate for this authority to monitor the task you completed?" and "How much did you respect the authority in the room?". We measure the *Authority Relation* using a three-measure scale, including the questions "How friendly was the authority in the room?", "How attentive was the authority to you?", and "How close did you feel to the authority in the room?"

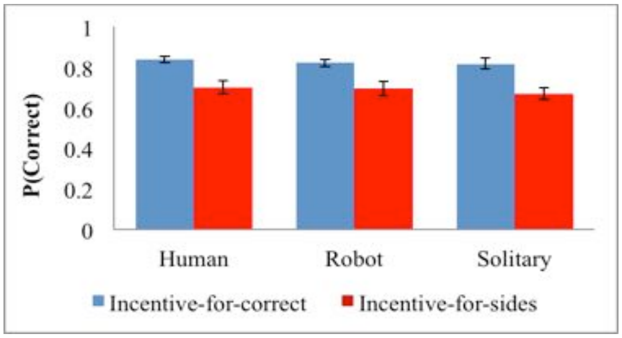

Fig 8. Percentage of correct-side identification in each block across conditions. In all conditions, accuracy was significantly higher in *incentive-for-correct* than in *incentive-for-sides*. Error bars show standard errors.

6.3 Monitoring

We measure the *Perceived Monitoring* of the human or the robot, compared to being alone, with two measures: A percentage scale labeled, "How much did the authority look at you as a percentage of total task time", and a 7-point measure asking "To what extent did you feel you were being monitored?"

6.4 Guilt

We measure the *Guilt* of the participant using a single question, "How guilty do you feel right now?"

6.5 Task Experience

We measure the participant's *Overall Experience* of the task, using a five-point Likert scale, asking how "clear", "easy", "enjoyable", and "interesting" the task was, "how the task felt to them" and "how attentive they were to the task".

7. RESULTS

To test Hypothesis 1, we calculated accuracy for each block, to see if people chose to provide false response to increase personal gain. In line with H1, participants were more accurate in identifying the side with more dots in incentive-for-correct trials than in incentive-for-side trials (we combined incentive-for-left and incentive-for-right trials, since no difference was found between those blocks). Fig. 8 shows the proportion of correct responses by condition and block. Repeated measures ANOVA revealed a significant effect for block type ($F(1,57) = 51.68$, $p < 0.001$), but there was no main effect for condition ($F(2,57) = 0.345$, $p = 0.71$), nor significant interaction between the two factors ($F(2,57) = 0.101$, $p = 0.9$). This pattern of results indicates that people cheated to some degree in each of the three conditions, since accuracy was markedly lower on incentive-for-side trials compared to incentive-for-correct trials, despite the fact that they were instructed to be as accurate as possible in all blocks.

To further examine if this reduction of accuracy in incentive-for-side represent cheating behavior, we calculated a "cheating index" for each participant. This index is the difference between the proportion of "beneficial errors" from the total number of trials (errors made to the high-paying side; e.g., errors to the left in *incentive-to-left* block) and that of "detrimental errors" (errors made to the low-paying side; e.g., errors to the right in *incentive-for-left* block):

$$CI = P\text{ (beneficial errors)} - P\text{ (detrimental errors)}$$

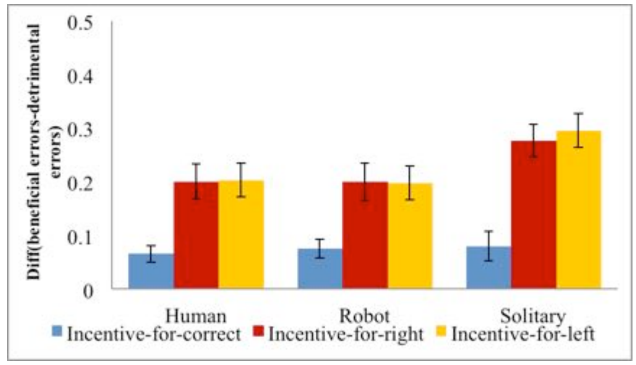

Fig. 9. Cheating Index was higher for incentive-for-side blocks than for incentive-for-correct block in all three conditions. When only considering incentive-for-side blocks, the cheating index in the solitary condition was higher than in the robot and human conditions. Error bars show standard errors.

If people try to cheat to increase personal gain, we would expect the proportion of errors to be biased toward the high-paying side. Thus, a higher CI indicates a higher level of cheating.

In line with this assumption, the averaged cheating index was 0.07 for incentive-for-correct block and 0.228 for incentive-for-side blocks ($F(1,57) = 34.381$, $P<0.001$). In addition, when only considering incentive-for-side blocks, the cheating index in the solitary condition ($M_{solitary} = 0.286$) was higher than in either the robot and human conditions ($M_{robot} = 0.199$ and $M_{human} = 0.201$). Post hoc analysis that compared the solitary condition to the two other conditions combined revealed that this difference is significant ($t=1.675$, $P=0.05$ one tailed).

In line with Hypothesis 2, we found that participants perceived the robot as the presence of an authority similarly to the way they perceived the human experimenter ($M_{robot} = 2.70$ and $M_{human} = 2.50$; $t(38) = 0.363$, $p=0.718$). However, Hypothesis 2a was not supported, as we found no significant difference between acceptance of the robot ($M_{robot} = 4.85$) and the human ($M_{human} = 4.30$) as authority ($t(38) = 0.984$, $p=0.3318$). In line with Hypothesis 2b, participants reported that they felt less related to a robotic authority than to the human authority, ($M_{robot} = 5.30$ versus $M_{human} = 6.00$), and expressed less respect to the robot ($M_{robot} = 4.60$ and $M_{human} = 5.45$), but in both cases this difference was not significant ($t(38) = 1.606$, $p=0.12$ and $t(38) = 1.643$, $p=0.11$, respectively).

Hypothesis 3 was only partially supported, since despite the fact that the human experimenter and the robot looked at the participants using the same algorithm, participants reported that they sensed being more monitored with a robotic social presence ($M_{robot} = 3.05$) than with a human presence ($M_{human} = 2.40$). However, this difference was not significant ($t(38) = 1.269$, $p=0.106$). In a similar vein, participants reported that they felt the robot authority looked at them for a longer period of time ($M_{robot} = 45.83\%$ of the time) than the human authority ($M_{human} = 26.89\%$). Independent-samples t-test revealed that this difference was significant ($t(38) = 2.567$, $p=0.02$).

As suggested by Hypothesis 4, people felt more guilty after dishonest behavior with a presence of a human in the room ($M_{human} = 2.42$) than when they are alone ($M_{solitary} = 2.20$). Surprisingly, people felt least guilty after dishonest behavior with a robotic social presence ($M_{robot} = 1.50$). While the overall effect was not significant ($F(2, 58) = 2.181$, $p=0.122$), planned contrast revealed that the difference in guilt between the robot and human conditions was significant ($t(38) = 1.99$, $p=0.05$). The difference between the robot and the solitary condition, however, was not significant. Thus, H4 was only partially supported.

Finally, to test Hypothesis 5, we calculated an overall experience grade for each participant based on the composite scale described in Section 6.5. The internal consistency was found to be high and acceptable (α Cronbach = 0.763). While experience was highest in the solitary condition ($M_{solitary}$=6.05), it was lowest in the human condition (M_{human}=5.65) and the robot condition was in between (M_{robot} = 5.99). One-way ANOVA revealed that these differences between conditions were not significant ($F(2, 57) = 1.023$, $p=0.33$). Thus Hypothesis 5 was not supported.

8. DISCUSSION

In our study, we found that both a human and a robot cause a similar reduction in cheating, by a significant amount compared to a person being alone in the room. We note that this effect transpired even though neither the human nor the robot seemed to be directly monitoring the person. We further did not find that the robot was perceived differently from the human experimenter as a presence of authority, and that people might be similarly accepting of the robot as an authority.

That said, they related to the robot and respected it as an authority slightly less when compared to a human. These two findings were trends, but did not yield significant results. In addition, participants felt significantly less guilty after they were dishonest with a robot as opposed to a human experimenter.

This leads us to suggest that social robots could be useful for monitoring tasks. Social and assistive robots could be used successfully to monitor task processes such as delivery of items, checking coats or returning car keys at valet stations, or could be use peripherally for monitoring when they perform other duties.

Based on our findings, these robots could be successful in promoting honesty, but might not be well-respected by humans. The results of our experiment indicate that we will need to design robots to create trust and rapport, and to make sure that they are viewed as a positive authority.

We controlled robot and experimenter gaze at the participant, but the robot was perceived as somewhat more of a monitoring presence. This is interesting given prior studies on how simple design features like the presence or absence of eyes and direction of gaze can drastically affect liking, trust, rapport, and willingness to cooperate with a robot [17, 31]. More research is needed to understand the effect of particular design features such as facial features, gaze, speech, and motion on the perception of being monitored.

We found a slight trend showing that participants enjoyed the experience most when they were alone and with the robot, compared to when they were with the experimenter, which they enjoyed less. This could be related to the fact that they felt less guilty about cheating with the robot. It could also be that the robot, being an interesting or novel device, piqued their interest and caused them to enjoy the task more, even though they felt monitored to the extent of cheating less (which we take to be a negative experience). The overall improvement in enjoyment could also, in turn, account for the lower guilt.

Finally, it is important to note that the effect of a robot's presence on people's honesty will clearly depend on people's increasing first-hand experience with robots' capabilities. For example, if people learn that robots monitor, record, and report their behavior, the robots' effect as honesty-evoking agents might increase. On the other hand, if robots will be deployed as a social presence only in order to discourage cheating, people will likely discover that fact and eventually ignore the robot's presence.

9. CONCLUSION

In this paper, we described the design of a new social robotic head to study the relationship between a robot's presence, design, and behavior, and human honesty. We present an interim prototype for the head and an experimental study evaluating whether the robot's social presence causes people to cheat less.

We found that a robot and a human similarly decrease cheating, but while not being perceived differently as an authority, they may be related-to and respected differently as such. We also find a trend for people's lower levels of guilt when cheating while being monitored by a robot.

That said, these are mere initial steps in our research path. We intend to expand this project by running the study with the fully constructed robotic head, enabling us to compare various designs for the head, face, and eyes. We will also study different behaviors and their effects on human honesty. Furthermore, we will mount the head on the mobile base to learn about the effects of robotic movement, proxemics, gestures on honesty.

Still, our results point to important implications for robots in the workforce, in education, and in public service settings, three environments in which honesty is key. Even with minimal design, suggesting mostly presence and gaze behavior, a robot was as successful as a human in decreasing cheating for money. This suggests that organizations and policy makers might consider the use of robots to monitor and supervise people in an effort to curb costly dishonest behavior.

10. ACKNOWLEDGEMENTS

We would like to thank Roberto Aimi for his work on the construction of the robotic head. This work was funded in part by a European Union FP7 Marie Curie CIG #293733 and by the National Science Foundation (IIS-0905148 & IIS-11165334).

11. REFERENCES

[1] Aroca, R. V, Péricles, A., de Oliveira, B.S., Marcos, L. and Gonçalves, G. 2012. Towards smarter robots with smartphones. *5th Workshop in Applied Robotics and Automation, Robocontrol*.

[2] Bainbridge, W., Hart, J., Kim, E. and Scassellati, B. 2008. The effect of presence on human-robot interaction. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2008)*.

[3] Bandiera, O., Barankay, I. and Rasul, I. 2009. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*. 77, 4, 1047–1094.

[4] Bartneck, C., Verbunt, M., Mubin, O. and Al Mahmud, A. 2007. To kill a mockingbird robot. *Proceeding of the ACM/IEEE international conference on Human-robot interaction - HRI '07* 81.

[5] Bateson, M., Nettle, D. and Roberts, G. 2006. Cues of being watched enhance cooperation in a real-world setting. *Biology letters*. 2, 3, 412–4.

[6] Bauer, A., Wollherr, D. and Buss, M. 2008. Human–robot collaboration: a survey. *International Journal of Humanoid Robots*.

[7] Bazerman, M.H. and Tenbrunsel, A.E. 2011. *Blind spots: Why we fail to do what's right and what to do about it*. Princeton University Press.

[8] Bhattacharjee, S., Gopal, R. and Sanders, G. 2003. Digital music and online sharing: software piracy 2.0? *Communications of the ACM*. 46, 107–111.

[9] BossaNova Robotics: *http://www.bnrobotics.com/*. Accessed: 2014-10-03.

[10] Burke, J., Coovert, M., Murphy, R., Riley, J. and Rogers, E. 2006. Human-Robot Factors: Robots in the Workplace. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 870–874.

[11] Canner, E. 2008. Sex, Lies and Pharmaceuticals: The Making of an Investigative Documentary about `Female Sexual Dysfunction'. *Feminism & Psychology*.

[12] Cialdini, R.B., Reno, R.R. and Kallgren, C.A. 1990. A focus theory of normative conduct: recycling the concept of norms to reduce littering in public places. *Journal of personality and social psychology*. 58, 6, 1015.

[13] Covey, M.K., Saladin, S. and Killen, P.J. 1989. Self-Monitoring, Surveillance, and Incentive Effects on Cheating. *The Journal of Social Psychology*. 129, 5, 673–679.

[14] Crocker, K.J. and Morgan, J. 1998. Is Honesty the Best Policy? Curtailing Insurance Fraud Through Optimal Incentive Contracts. *J of Political Economy*. 106, 355.

[15] DePaulo, B.M. and Kashy, D.A. Everyday lies in close and casual relationships.

[16] Diener, E., Fraser, S.C., Beaman, A.L. and Kelem, R.T. 1976. Effects of deindividuation variables on stealing among Halloween trick-or-treaters. *Journal of Personality and Social Psychology*. 33, 2, 178.

[17] DiSalvo, C.F., Gemperle, F., Forlizzi, J. and Kiesler, S. 2002. All robots are not created equal: the design and perception of humanoid robot heads. *Proc of the 4th conference on designing interactive systems (DIS2002)* 321–326.

[18] Gino, F., Ayal, S. and Ariely, D. 2009. Contagion and differentiation in unethical behavior: the effect of one bad apple on the barrel. *Psychological science*. 20, 3, 393–8.

[19] Gino, F., Norton, M.I. and Ariely, D. 2010. The counterfeit self: the deceptive costs of faking it. *Psychological science*. 21, 5, 712–20.

[20] Groom, V., Chen, J., Johnson, T., Kara, F.A. and Nass, C. 2010. Critic, compatriot, or chump?: Responses to robot blame attribution. *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI'10)*.

[21] Hamblin, R.L., Hathaway, C. and Wodarski, J.S. 1971. Group contingencies, peer tutoring, and accelerating academic achievement. *A new direction for education: Behavior analysis*. 1, 41–53.

[22] Hoffman, G. 2012. Dumb Robots , Smart Phones : a Case Study of Music Listening Companionship. *RO-MAN 2012 - The IEEE Int'l Symposium on Robot and Human Interactive Communication* 358–363.

[23] Hoffman, G. and Ju, W. 2014. Designing Robots With Movement in Mind. *Journal of Human-Robot Interaction*. 3, 1, 89.

[24] Ju, W. and Takayama, L. 2011. Should robots or people do these jobs? A survey of robotics experts and non-experts about which jobs robots should do. *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems* 2452–2459.

[25] Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H. 2004. Interactive robots as social partners and peer tutors for children: A field trial. *Human-Computer Interaction*. 19, 61–84.

[26] Kaniarasu, P. and Steinfeld, A. 2014. Effects of blame on trust in human robot interaction. *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN'14)*.

[27] Lee, K.M., Peng, W., Jin, S.-A. and Yan, C. 2006. Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human-Robot Interaction. *Journal of Communication*. 56, 4, 754–772.

[28] Mas, A. and Moretti, E. 2006. *Peers at work*.

[29] Mazar, N., Amir, O. and Ariely, D. 2008. The dishonesty of honest people: A theory of self-concept maintenance. *Journal of marketing research*. 45, 6, 633–644.

[30] Murphy, K.R. 1993. *Honesty in the workplace.* Thomson Brooks/Cole Publishing Co.

[31] Mutlu, B., Forlizzi, J. and Hodgins, J. 2006. A storytelling robot: Modeling and evaluation of human-like gaze behavior. *Humanoid Robots, 2006 6th IEEE-RAS International Conference on* 518–523.

[32] Nagin, D., Rebitzer, J., Sanders, S. and Taylor, L. 2002. *Monitoring, Motivation and Management: The Determinants of Opportunistic Behavior in a Field Experiment*.

[33] Reno, R.R., Cialdini, R.B. and Kallgren, C.A. 1993. The transsituational influence of social norms. *Journal of personality and social psychology*. 64, 1, 104.

[34] Short, E., Hart, J., Vu, M. and Scassellati, B. 2010. No fair!! An interaction with a cheating robot. *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI'10)*.

[35] Takayama, L., Ju, W. and Nass, C. 2008. Beyond Dirty, Dangerous and Dull : What Everyday People Think Robots Should Do. *HRI '08: Proceeding of the ACM/IEEE international conference on Human-robot interaction*.

[36] Tanaka, F., Cicourel, A. and Movellan, J.R. 2007. Socialization between toddlers and robots at an early childhood education center. *Proceedings of the National Academy of Sciences of the United States of America*. 104, 46, 17954–8.

[37] Tanaka, F. and Ghosh, M. 2011. The implementation of care-receiving robot at an English learning school for children. *Human-Robot Interaction (HRI), 2011 6th ACM/IEEE International Conference on* 265–266.

[38] Vazquez, M., May, A., Steinfeld, A. and Chen, W.-H. 2011. A deceptive robot referee in a multiplayer gaming environment. *2011 International Conference on Collaboration Technologies and Systems (CTS)* 204–211.