Studying the Opposing Effects of Robot Presence on Human Corruption

Michal Roizman*, Guy Hoffman*, Shahar Ayal†, Guy Hochman†, Michal Reifen Tagar† and Yossi Maaravi‡
*Media Innovation Lab, †School of Psychology, ‡Adelson School of Entrepreneurship (ASE)
Interdisciplinary Center (IDC) Herzliya, Herzliya 46150, Israel
Email: {Michal.roizman,hoffman,ayal,ghochman,Michal.Reifen,myossi}@idc.ac.il

***Abstract*—Social presence has two opposing effects on human corruption: the collaborative and contagious nature of another person's presence can cause people to behave in a more corrupt manner. In contrast, the monitoring nature of another person's presence can decrease corruption. We hypothesize that a robot's presence can provide the best of both worlds: Decreasing corruption by providing a monitoring presence, without increasing it by collusion. We describe an experimental study currently underway that examines this hypothesis, and report on initial findings from pilot runs of our experimental protocol.**

Fig. 1. The peep-hole die cups used in the experiments. Participants privately roll the dice and decide whether to report honestly on the result.

I. Introduction

Corruption is a broad term used to describe varying degrees of human immorality. While usually associated with large-scale fraud, the term can also be applied to one's day-to-day decisions, where small actions such as downloading music illegally, taking office supplies from work, or inflating expense report items are committed by people generally perceiving themselves as "good" [1].

Such decisions are influenced not just by individual differences and societal norms, but often by specific circumstances. A sense of being monitored by others can decrease immoral behavior, even with minimal cues [2]. On the other hand, collaborative settings can increase corruption. A recent study found that when there are signs of corruption in one's surrounding, an individual tends to act in a similar manner. Furthermore, people were found to favor collaboration over moral standards, and cheat more when they feel they benefit a group and not just themselves [3]. In other words, a social presence can work both to decrease and to increase immoral behavior.

As robots become more prevalent and integrate further into society, their presence around conflicting situations where a moral decision has to be determined is inevitable. A recent study examined the effect of robot presence on human moral behavior, and has concluded that participants' cheating decreases by equal measures when there is a robot in the room as with a human present, in comparison to working alone [1]. This indicates that a robot can contribute to the corruption-mitigating monitoring aspects of social presence. An open question remains regarding the effects of a robot's presence in a collaborative setting in which cheating is an option.

This paper describes our work in progress studying the effect of a robotic collaborator on human moral decisions. By controlling for both social presence and corrupt collaborative cues, we aim to tease apart which mechanisms are in play when humans and robotics collaborate in a morally ambiguous situation.

In our study we compare people's moral decisions when they are alone, when they are collaborating with a human, and when they are collaborating with a robot. We also compare situations with and without corruption signals in the collaboration. We use an established protocol from behavioral psychology, adding a condition of collaborating with a small non-anthropomorphic robot. We report on our hypotheses, our study protocol, and initial qualitative findings.

II. Related Work

Cheating behaviors are a major social and personal issue, estimated to cost billions of dollars annually [1]. Recent studies find collaboration to encourage corruption. For example, Weisel and Shalvi showed that people favored collaboration over being true to moral standards [3].

Robots are often seen as entities with some elements of "theory of mind", including morality [4]; however, people view robots differently than humans when it comes to moral decisions [5]. Still, robots have been shown to influence human moral behavior, as shown in a study that found that people's cheating decreases equally with a robot in the room as with a human present, compared to when they were alone [1].

This suggests that robots have monitoring-related effects on corruption. An open question remains how robots effect the collusion-related aspects of corruption. Addressing this question is the aim of the study described in this paper.

III. Experimental Design

To tackle this question, we designed a human-robot interaction protocol based on the "die under cup" paradigm used in Weisel and Shalvi [3], and drawing on Krach *et al.*'s deception

978-1-4673-8370-7/16/$31.00 © 2016 IEEE

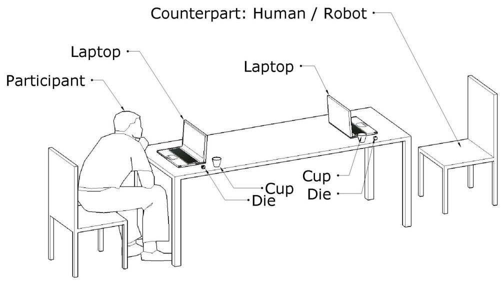

Fig. 2. Physical setup of the experiment. In the COMPUTER condition, people play vis-a-vis a laptop; in the HUMAN condition, vis-a-vis a human confederate; in the ROBOT condition, a robot. In all cases, participants are in fact facing a fixed dice-rolling series generated by their own laptop.

setup [6]. In the latter, participants believed that they played a game of iterated prisoner's dilemma with two robots, a human, and a laptop. In fact, they actually faced the same opponent algorithm unrelated to their perceived opponent.

Since we want to both compare the type of social presence (computer, robot, and human), and the corruptive signaling effects, we use a 3 (type of presence—between subjects) x 2 (corruption signaling—within subject) mixed design.

IV. Method

Participants play a dice game for money with one of three of counterparts: Computer, human or robot, in a between-subject manner. Participants enter a room and are seated in front of a table with a standard six-sided die, a cup with a peephole (Fig. 1) and a computer with the experiment software. In the COMPUTER condition, the participants play with an additional computer as their counterpart. In the HUMAN and ROBOT conditions, they face either a human or a robot on the other side of the table (Fig. 2). Each participant privately rolls a die using the cup, looks at the result through the peephole, and reports the outcome in the computer software.

In each round, the counterpart rolls the die first and then reports the outcome to the system. This number then appears on the participant's computer. Then the participant rolls their own die and reports an outcome as well. When both participant and counterpart report a same number ("double"), they both earn an amount of money equivalent to the number they reported multiplied by the equivalent of circa $0.05 (e.g. a double of 6 will earn both counterparts circa $0.30). In case the numbers differ, they earn nothing.

In reality, each participant sees the same sequence of numbers as their counterpart's rolls. In the first block of 20 rounds (NO-CORRUPTION), these are uniformly distributed numbers between 1–6. In the second found (CORRUPTION), there are only 5s and 6s.

A. Research Hypotheses

There are two opposing effects of social presence on corruption: monitoring decreases corruption; collusion increases it. We hypothesize that the robot will provide the best of both worlds: Decreasing corruption by providing a monitoring presence, without increasing corruption by way of collusion. We therefore hypothesize that:

1) In the HUMAN condition, there will be less corruption than the COMPUTER condition in the NO-CORRUPTION condition, but more in the CORRUPTION condition.
2) In the ROBOT condition, there will be less corruption than the COMPUTER condition in both conditions.

V. Pilots and Preliminary Findings

We ran eight pilot studies so far to test our protocol, with student volunteers participating in the game for real compensation, and then being interviewed about the interaction.

The pilots provided several insights, the most salient of which were: (a) Participants emphasized being in an academic setting. For some this meant they shouldn't cheat for fear of being disciplined. Others mentioned, perhaps jokingly, that it felt fair to be recouping some of their tuition. This suggests perhaps not using a student sample in this study. (b) The amount of money matters. Several participants commented that they would not compromise their integrity for such small amounts of money. Howevr, previous research suggests that small rewards facilitate cheating [7]. We are conducting further pilots to establish the optimal monetary reward.

VI. Conclusion

We present work in progress toward an experimental study aimed to tease apart the differences in moral decisions when collaborating with a computer, a human, and a robot. The novel contribution of this work is to understand how robots play into the opposing effects of social presence on corruption, and whether robots could provide a "best of both worlds" solution that can increase people's morality and ethical decisions.

Acknowledgment

We would like to thank Yahav Amsalem and Noam Fisher for their assistance in designing and running the studies.

References

[1] G. Hoffman, J. Forlizzi, S. Ayal, A. Steinfeld, J. Antanitis, G. Hochman, E. Hochendoner, and J. Finkenaur, "Robot Presence and Human Honesty: Experimental Evidence," in *Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI'15)*. Portland, OR: ACM, 2015.

[2] M. Bateson, D. Nettle, and G. Roberts, "Cues of being watched enhance cooperation in a real-world setting." *Biology letters*, vol. 2, no. 3, pp. 412–4, Sep. 2006.

[3] O. Weisel and S. Shalvi, "The collaborative roots of corruption," *Proceedings of the National Academy of Sciences*, vol. 112, no. 34, pp. 10 651–10 656, 2015.

[4] H. M. Gray, K. Gray, and D. M. Wegner, "Dimensions of mind perception," *Science*, vol. 315, no. 5812, p. 619, 2007.

[5] B. F. Malle, M. Scheutz, and J. Voiklis, "Sacrifice one for the good of many? people apply different moral norms to human and robot agents," in *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2015, pp. 117–124.

[6] S. Krach, F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher, "Can machines think? Interaction and perspective taking with robots investigated via fMRI." *PloS one*, vol. 3, no. 7, p. e2597, Jan. 2008.

[7] L. K. John, G. Loewenstein, and S. I. Rick, "Cheating more for less: Upward social comparisons motivate the poorly compensated to cheat," *Organizational Behavior and Human Decision Processes*, vol. 123, no. 2, pp. 101–109, 2014.