Monetary-Incentive Competition between Humans and Robots: Experimental Results

Alap Kshirsagar*, Bnaya Dreyfuss†, Guy Ishai†, Ori Heffetz†* and Guy Hoffman*
*Cornell University, Ithaca, New York, USA
†The Hebrew University of Jerusalem, Jerusalem, Israel
Email: ak2458@cornell.edu, benaya.dreyfuss@mail.huji.ac.il, guy.ishai@mail.huji.ac.il, oh33@cornell.edu, hoffman@cornell.edu

***Abstract*—In a controlled experiment, participants ($n = 60$) competed in a monotonous task with an autonomous robot for real monetary incentives. For each participant, we manipulated the robot's performance and the monetary incentive level across ten rounds. In each round, a participant's performance compared to the robot's would affect their odds in a lottery for the monetary prize. Standard economic theory predicts that people's effort will increase with prize value. Furthermore, recent work in behavioral economics predicts that there will also be a discouragement effect, with stronger robot performance discouraging human effort, and that this effect will increase with prize. We were not able to detect a meaningful effect of monetary prize, but we found a small discouragement effect, with human effort decreasing with increased robot performance, significant at the $p < 0.005$ level. Using per-round subjective indicators, we also found a positive effect of robot performance on its perceived competence, a negative effect on the participants' liking of the robot, and a negative effect on the participants' own competence, all at $p < 0.0001$. These findings shed light on how people may exert work effort and perceive robotic competitors in a human-robot workforce, and could have implications on labor supply decisions and the design of compensation schemes in the workplace.**

***Index Terms*—Human-Robot Competition; Reference-Dependent Preferences; Loss Aversion; Perceived Competence**

I. Introduction

In this paper, we present an experiment studying the effort of people competing for a real monetary prize with a robot, and how the robot's performance affects people's effort and their attitude toward the robot. People compete with each other in workplaces, politics, sports, and other contexts, both for monetary and non-monetary gains. When humans make decisions on whether and how much to invest in a competitive task, they can be motivated by several factors. Traditional economic models of decision-making often focus exclusively on the direct value of the reward for the task. However, recent work in behavioral economics highlights additional considerations, in particular, that the psychological value of winning a reward can be affected by prior expectations. These motivations are not just important to the individual, but also to society at large as they affect labor supply decisions and the design of compensation schemes in the workplace.

This project was supported in part by the Planning and Budgeting Committee and the Israel Science Foundation (grant no. 1821/12). The authors thank David Gill and Matthew Rabin for their helpful comments.

Behavioral economics researchers have found evidence supporting the notion that people are "loss-averse," i.e., perceive the disappointment resulting from the loss of a reward more strongly than the satisfaction resulting from gaining an equivalent reward. These perceived gains and losses, however, are not objectively determined but rather depend on a subjective reference point. In a competition, the reference point often depends on people's own efforts and that of their competitors.

With robots entering the workforce, and robots along with other artificial intelligence (AI) agents playing an increasing part in the economy, we can expect situations where people's real economic outcomes depend on a robot's performance. Also, people may compete with their robotic co-workers, e.g., to win resources or to demonstrate capabilities. We are therefore interested in understanding how people's effort is affected by a robot's performance in a competitive situation.

This work is also motivated by the dearth of Human-Robot Interaction (HRI) research that focuses on human-robot competition, especially outside a game environment. Beyond the study of people's competitive effort, we are therefore also interested in understanding human attitudes towards the robot and towards themselves in competitive settings, and how those are affected by the robot's performance.

II. Background

A. Human-Robot Competition

A large segment of the HRI literature is concerned with collaboration between humans and robots, where both agents share the same goal [1]–[3]. Competitive scenarios, in contrast, have rarely been explored.

In the AI community, competitive games have served as a benchmark for a variety of algorithms, including Checkers [4], Backgammon [5], Chess [6], and more recently, Go [7]. However, the structure of these games was used to illustrate the learning, reasoning, and planning capacities of the algorithm, and was not focused on the human's effort and attitudes in these competitions. In addition, engaging and complex games are not good proxies for the rote work-like task competitions we are concerned with in this work.

HRI research, even in game-playing scenarios, has mostly looked at collaborative games, where the direct competition with the robot is not emphasized [8], [9]. There are a few notable exceptions: Mutlu et al. [10] explored the perceptions

978-1-5386-8555-6/19/$31.00 ©2019 IEEE

of an ASIMO robot in competitive and cooperative scenarios, and found that male participants were more engaged with the task when competing with the robot but found it less socially desirable than a cooperative robot. Short et al. [11] studied how a cheating robot influences attributions of mental state and intentionality in a human-robot competition, and found increased engagement and mental state attribution when the robot cheated as compared to the control condition in which the robot played fairly. In both cases, the study focused on a playful game competition. Neither considered a human-robot competition involving repetitive or mundane tasks for a real monetary reward. The work presented here thus provides new insights into a so-far unexplored area of HRI research, namely a real-effort competition between a human and a robot for monetary incentives.

B. Expectations-Based Loss Aversion

The main economic theory we are testing in this experiment is expectations-based reference dependent (EBRD) loss aversion. When making economic decisions, humans have long been shown to be risk-averse, tending to lower uncertainty even at the cost of lowering expected payoff (for a recent review, see [12]). Traditionally, this was associated with a concave utility function [13], where utility increases at a diminishing rate. However, more recent work has shown that a concave utility function cannot plausibly explain both small- and large-stakes risk aversion [14]. One notable alternative explanation is loss aversion, in which losses are weighted more heavily than gains—both defined relative to a reference point. This creates extra local concavity of the utility function around the reference point, which, in addition to the traditional (reference-independent) concavity, can explain both small- and large-scale risk aversion. The term "loss aversion" was coined and first presented by Kahneman and Tversky [15], and became increasingly common among economists.

Within this theoretical framework, EBRD utility models were developed with a special focus on how the reference point is determined. These models posit that the reference point depends on people's expectations. Notable early EBRD models include [16]–[18]. More recently Kőszegi and Rabin [19]–[21] (henceforth KR) developed a more comprehensive version of an EBRD model, which became popular for economic modeling of loss aversion. The KR model adds to the traditional model two main features: the first relates to how people react to departures in their consumption from the reference point, which can be either a deterministic outcome or a distribution of outcomes. The second feature is a formation of the reference point according to a person's rational expectations.

The model has been experimentally evaluated in a variety of contexts, including a competition between two people. We draw much of our experimental design and theoretical predictions from one of these experiments, conducted in 2009 by Gill and Prowse [22] (henceforth GP). In GP, two people, called "First Mover" and "Second Mover" completed a computerized real-effort task sequentially, and the probability of winning a prize was determined by the effort exerted in the task relative to the competitor. Consistently with EBRD predictions, GP found an interacted *discouragement effect*, in which the Second Mover's effort was negatively affected by the First Mover's effort, increasingly with prize value ($p = 0.04$).

Other studies investigating EBRD predictions focused on labor supply decisions under uncertainty and on people's attachment to objects which they expect to possess, also known as the *endowment effect*. These studies provide some supportive evidence, although its generalizability, robustness, and interpretation are still actively debated and scrutinized (for recent examples and detailed discussions of earlier work, see, e.g., [23] and [24]).

Our experiment extends existing empirical evidence in several ways. First, it includes a simultaneous, rather than GP's sequential, competition. Second, it replaces GP's human-human competition with a human-robot one. Third, we measure people' attitudes towards robots in a competitive situation where money is at stake. Finally, we have added several design features in order to make the theoretical predictions robust to a more general version of the EBRD model [21].

III. Competition Design

We developed a within-subject experimental protocol to study human-robot interaction in a monotonous competitive environment. Participants competed with a robot on a repetitive task that involved counting letters and moving a block in the workspace.[1] We manipulated two independent variables: robot performance and monetary prize level.

A. Letter Counting and Block Placement Task

The participant and the robot each receive a randomly generated string of 20 characters. They have to count the number of 'G' letters in their texts and place a block in the corresponding bin. There are either 3, 4 or 5 'G' letters in the text and three bins marked with '3', '4' and '5' in the workspace. After placing the block in a bin the participant has to click a button on the screen to validate their counted number. Following a correct placement, the participant gets one point and receives the next string of characters. For an incorrect placement, they do not receive a point and the submission of block placement gets disabled for 10 seconds. At that stage, they can still re-count the letters in the text and move the block around, but cannot click the button to validate. After 10 seconds, the participant can again submit their block placement for verification. Each competition round lasts two minutes. However, the participant does not have to compete for a full two minutes and can choose to stop working at any time.

B. Prize Scheme

The participant's chance of winning the prize for each round depends upon the difference between the robot's final score (denoted e_r for "robot effort") and participant's final score (denoted e_h for "human effort"). If the scores are the same,

[1] All experimental materials, including full instructions, questionnaires, screen-shots, recruiting materials, video recordings of the setup, and the resulting data, are available at www.nber.org/~heffetz.

the participant has a 50% chance of winning the prize. The probability of winning the prize (p) increases/decreases by 1% per unit difference in points, as given in Eq.1.

$$p = \frac{e_h - e_r + 50}{100} \tag{1}$$

If the full two-minute round elapses, e_h and e_r are the actual number of tasks completed. If a participant chooses to stop the round early, e_r will be equal to the robot's projected score, which is equivalent to assuming that the robot would have continued uninterrupted until the end of the two-minute round; e_h will be the participant's score at the time of stopping. The projected score is calculated based on the average speed, as $e_r(final) = \frac{e_r(now)}{time(now)} \times 120$.

C. Robot and User Interface

We use an off-the-shelf WidowX Mark II robot arm to perform the letter counting and block placement task. The experiment setup is shown in Fig. 1. An Orbbec Astra vision sensor detects the block position using the `cmvision` package [25], which is fed to the motion planner *MoveIt!* [26]. Within *MoveIt!* we use the inverse kinematics solver `trac-IK` [27] to generate joint trajectories for the robot arm. Due to the vision sensor feedback, the robot is able to recover from failures and picks up the block from its correct position to place it in the correct bin. We also use the input from the robot's vision sensor to verify the human's block placement for score-keeping and to trigger the 10-second penalties.

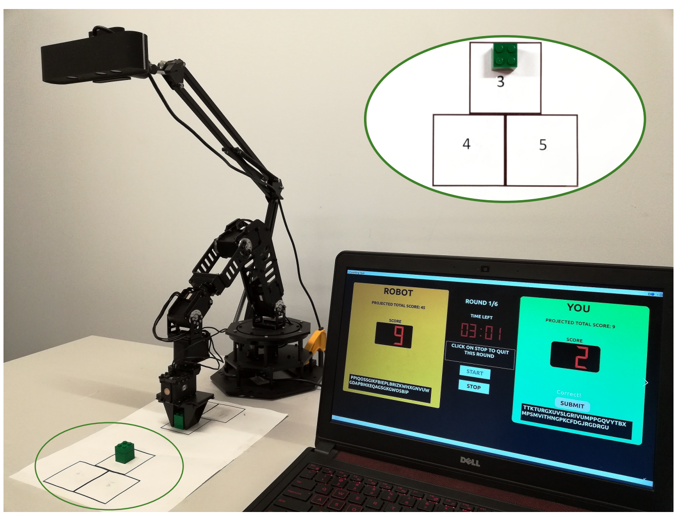


Fig. 1: Experiment Setup

The user interface, as shown in Fig.2, displays the points accumulated so far ('Score'), the total points expected at the end of the two-minute round ('Projected Final Score'), the monetary prize for the round, the time left in the round and the participant's chance of winning the prize at each instant ('Your Chance of Winning the Prize, if You Stop Now'). The screen also has buttons for starting and stopping the round, and for submitting the block arrangement for verification ('Submit'). Based on pilot studies in which participants noted that they did not have a sense of their winning odds, the current probability of winning is additionally read out by a robotic voice every five seconds. At the end of each round, another screen displays the final scores and participant's probability of winning the prize for that round. Participants use a USB mouse to interact with the screen. The entire software, including the robot's motion controller, computer vision and user interface, is built within the Robot Operating System framework (ROS) [28].

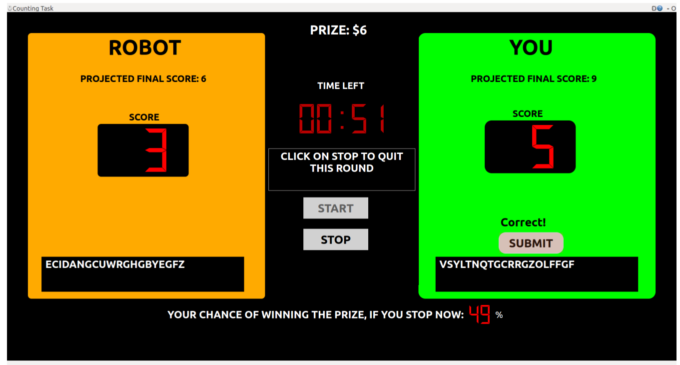


Fig. 2: User Interface Screen

D. Lottery

The winner of each round is decided in a lottery based on the percentage chance of winning at the end of the round. We use a public website for this purpose.[2] For each paying round, we roll a fair 100-sided die on that website. If the die-roll result is less than or equal to the participant's chance of winning the prize, the participant wins the prize for that round.

IV. Theoretical Predictions

Our basic theoretical predictions are identical to GP's. We present a brief, intuitive derivation. For the full mathematical derivation of the predictions, see GP [22].

According to economic theory, a human's choices maximize a utility function. In our context, the utility is a function of the human's effort, e_h. Traditional models of labor supply highlight two utility terms. The first, $-C(e_h)$, is an increasing cost function of effort. The second is the expected payoff, which we can model in our scenario as pv, with p the probability of winning (Eq. 1), and v the size of the prize. KR's EBRD model includes two additional terms that represent expected gains and losses. With probability p, the human gets the prize v and experiences a gain relative to the potential outcome of not winning the prize; the gain equals $(1-p)v$. With probability $1-p$ the human gets nothing and experiences a loss relative to the potential outcome of winning the prize; the loss equals $-\lambda pv$. The weighting of a loss relative to a gain by $\lambda > 1$ formalizes loss-aversion. The resulting utility function is the following:

$$U(e_h) = -C(e_h) + pv + p(1-p)v - (1-p)\lambda pv. \tag{2}$$

[2]http://www.roll-dice-online.com/

Therefore, standard economic theory, without reference dependence, predicts no impact of the robot's projected score on the human's chosen score. The maximization of only the first two terms means that the human simply wants to maximize their expected payoff without exerting too much effort. However, the optimal effort level in EBRD utility does depend on e_r; specifically, it decreases in e_r. To see why, notice that the last two terms constitute a quadratic function of p and an increasing linear function of v. The quadratic function is U-shaped since losses loom larger than gains ($\lambda > 1$). This means that an additional utility ΔU from a small addition to the human's effort Δe_h is worth less when the base probability p is smaller, i.e., when the robot's projected score is higher. Therefore, such EBRD models predict that a higher robot's projected score lowers the human's performance. Following past work, we refer to this as a 'discouragement effect'. The magnitude of this effect should increase with the size of the prize (v). For formal proofs see GP [22].

V. Research Questions and Hypotheses

We were interested to know whether and to what extent a competing robot's performance and the monetary prize of the competition change the effort exerted by a human on a real-effort task. We followed GP and the derivation above to formulate the following hypotheses:

- **H1a** — The human's effort will increase with prize value.
- **H1b** — The human's effort will decrease with robot effort.
- **H1c** — The discouragement effect in **H1b** will increase with prize value.

H1a is a trivial result of increasing utility from money—a fundamental assumption in economic theory. It is therefore predicted by all economic models, traditional as well as EBRD. **H1b** is uniquely predicted by EBRD models. It was motivated in the previous section; it is formally proved with a weak inequality in GP Proposition 2. **H1c** is a special case of **H1b**. It is a direct result of **H1a**, **H1b**, additional technical regularity conditions, and a quadratic approximation of the effort cost function—as proved in GP Proposition 3. The intuition behind **H1c** is that the discouragement effect (**H1b**) is predicted only to the extent that subjects are motivated to win the prize (**H1a**).

In addition to participants' economic behavior, we were also interested to know whether and to what extent the robot's performance would affect participants' attitudes toward the robot and toward themselves. We trivially believed that people would consider a robot more competent the better it performed (**H2a**). We also thought that the robot's performance would affect their liking of the robot, although we did not have a directed hypothesis on whether people would like a better-performing robot more or less (**H2b**). Based on the anecdotal evidence found in several other works, e.g., [10], [29], we also hypothesized that people will be more self-critical when interacting with a better-performing robot (**H2c**). Our second set of hypotheses is thus:

- **H2a** — The human's rating of the robot's competence will increase with robot effort.
- **H2b** — The human's liking of the robot will be affected by robot effort (two-tailed).
- **H2c** — The human's rating of their own competence will decrease with robot effort.

Finally, we were interested to explore whether and how individual differences affect these outcomes. We did not have directional hypotheses, but measured the following individual traits in order to analyze their interaction with the above behaviors and attitudes:

- **Money Belief** — One's attitude towards money.
- **Self-Efficacy** — One's belief in their ability to succeed.
- **Competitiveness** — One's tendency to be competitive.

VI. Method

A. Procedure

Participants sat in front of the screen, facing the competitor robot. After signing the written consent form, they read the printed instructions. They then answered a comprehension quiz, designed to test their understanding of the prize scheme, participated in a demonstration of the lottery resolution procedure, and competed in a practice round to familiarize themselves with the task. Then the experimenter left the room and the main part of the experiment began. Participants completed 10 competition rounds. After each round, they filled a short questionnaire about these three points:

1) Robot Competence: "Please rate how much you consider the robot to be competent on the following scale:" (1–5)
2) Robot Likability: "Please rate how much you like the robot on the following scale:" (1–5)
3) Self Competence: "Please rate how true the following sentence is for you with respect to this task: I feel confident in my ability to do this word-counting task well." (1–7).[3]

We collected no additional questionnaire data per round beyond these three single item scales. The prizes for rounds were randomly drawn from [\$0.1, \$0.2,...,\$3.8]. This prize range was inspired by GP's, albeit in their case it was in UK pounds rather than in US dollars. The robot's speed was randomly chosen to correspond to a final robot score of one of [5,...,45] and was kept constant throughout the round. However, due to inaccuracies in the robot's motion planner, we ended up with a small fraction (2.24%) lower than 5 and another 0.45% at 46. At the end of 10 rounds, participants completed validated questionnaires on self-efficacy [32], money-belief [33], competitiveness [34]. Also, participants were asked to give written responses to these two questions:

- Please write a few sentences about your experience of competing with this robot.
- Finally, if you have any comments or thoughts you would like to share with us, please write them here. We are especially curious to know: how did you decide in each round how strongly to compete?

[3]We followed the Godspeed questionnaire [30], from which we adapted Q1&2, using a 5-point scale, and the Self Competence questionnaire [31], from which we adapted Q3, using a 7-point scale.

TABLE I: Summary of Robot and Human Efforts

Round	Mean R	SD R	Mean H	SD H	Min R	Min H	Max R	Max H	Mean Inc	SD Inc	Min Inc	Max Inc
Practice	23.97	1.72	15.78	5.30	14	0	26	24	1.93	1.27	0	6
1	24.93	14.33	19.67	3.73	3	11	45	27	2.15	1.44	0	6
2	26.10	15.13	20.18	4.65	5	8	46	33	2.18	1.30	0	5
3	24.52	15.60	19.77	4.37	1	11	46	31	2.45	1.50	0	6
4	25.35	13.75	19.75	6.02	5	0	44	31	2.27	1.44	0	6
5	22.97	14.84	20.15	5.38	4	0	45	32	2.37	1.31	0	6
6	22.32	14.51	19.98	7.13	4	0	45	34	2.43	1.39	0	6
7	23.40	14.98	20.72	6.70	3	0	45	33	2.12	1.63	0	7
8	25.85	15.09	20.63	7.50	3	0	46	35	2.27	1.33	0	5
9	26.62	14.90	21.42	5.55	3	5	45	31	2.10	1.42	0	5
10	22.07	15.19	21.97	6.87	5	0	45	32	2.25	1.46	0	6

SD denotes Standard Deviation, Min denotes Minimum, Max denotes Maximum, R and H denote, respectively, Robot and Human Scores, and Inc denotes the human's incorrect attempts.

We collected no questionnaire data per participant other than the questionnaires mentioned above and basic demographic information.

B. Participants

A total of 61 participants completed the study. Participants were recruited from an online pool of students. The invitation text described the experiment as "making decisions in the presence of a robotic arm", but did not mention a competition or prize money, only that "participants will receive $10 for their time". There was an error in the data logging system for one of the participants. Therefore, we do not include this participant in the data analysis. The valid set of 60 participants consisted of 43 females, 16 males and one participant of unspecified gender. One participant did not fill out the questionnaire after one of the rounds. We include this participant's responses for other rounds in the data analysis. We thus have $n = 600$ observations for most analyses, and $n = 599$ for the ones involving the per-round self-report scales.

Each study session lasted about 45 minutes. The participants received $10 as the show-up fee and any additional amount won in the competition. The average payment was $18.89, including the $10 show-up fee.

VII. Results

In response to the recent reproducibility crisis [35], and as part of an effort for more reproducible studies in HRI, we consider the p-value threshold for statistical significance in all of our results to be 0.005 instead of the more commonly used value of 0.05, as suggested by [36]. We therefore also use confidence intervals of 99.5% instead of 95% in our graphs. Furthermore, we will generally attempt to minimize using the term "significant," and instead report p-values directly. Finally, to err on the conservative side of our conclusions and to make our results easier to parse, we use two-tailed p-values even when we have directional hypotheses.

A. Summary of Behavior

Table I summarizes robot and human performance. We see a considerable variation in the participants' scores, ranging from 0 to 35. Overall, excluding the practice round, mean human score (20.42) is lower than mean robot score (24.41), and increases from 19.67 in the first round to 21.97 in the tenth round. The number of incorrect submissions per participant in a round ranged from 0 to 7; on average, in each round humans had two to three incorrect submissions.

B. Effect of Prize and Robot Effort on Human Effort

To test hypotheses **H1a–c**, we ran a fixed-effects multivariate regression controlling for round number and participant ID. Table II, which is a fixed-effects version of GP Table 2, reports our main-coefficient estimates. **H1a** and **H1b** predict the estimated average effect of prize and robot score to be positive and negative, respectively. **H1c** (a special case of **H1b**) predicts a negative estimate of the interaction term Robot Score × Prize.

TABLE II: Effects of Prize and Robot Effort on Human

Variable	Estimate	Std. Error	t-Ratio	p-Value
Prize	0.242	0.334	0.73	0.468
Robot Score	-0.049	0.028	-1.75	0.080
Robot Score × Prize	0.003	0.012	0.24	0.814

($n = 600$)

As Table II shows, **H1c** is not supported: the interaction coefficient is small, positive, and not statistically significant ($p = 0.814$; an F-test reveals that jointly, the three coefficients are significant, $p = 0.005$). Evaluating **H1a** and **H1b**, which concern average effects in Table II's interacted (and hence, nonlinear) specification, is less straightforward and requires additional calculations.[4] However, because the interaction coefficient's estimate is close to zero, and for ease of presentation, we instead proceed by assuming it to be exactly zero, and estimate a non-interacted specification that replicates Table II but drops the interaction term. In this non-interacted specification, reported in Table III, **H1a** and **H1b** are directly testable by looking at the estimated coefficients on Prize and Robot Score.

Table III shows that while **H1a** is not meaningfully supported ($p = 0.073$), **H1b** is supported ($p = 0.002$). Thus,

[4]Evaluating **H1a** and **H1b** from the interacted specification in Table II can be done using the Delta method for estimating the local effects of Prize and of Robot Score at different values, including the effects at the average, and the average effects. We found this analysis to yield almost identical coefficients as the more straightforward method that follows in the text, and thus to support the same conclusions. This is not surprising given that the interaction term is close to zero.

TABLE III: Average Effects of Prize and Robot Effort

Variable	Estimate	Std. Error	t-Ratio	p-Value
Prize	0.310	0.172	1.80	0.073
Robot Score	-0.043	0.014	-3.11	0.002

$(n = 600)$

while we find no statistically detectable reaction of human effort to monetary prize—and therefore no detectable support for a fundamental assumption of any economic model—we find that increasing the robot's score discourages the human from performing better at the task. To illustrate the size of the estimated effect: increasing the robot's score from 5 to 45 decreases the human's score by an average of 1.72, that is, by 8.4% relative to the average baseline. For comparison, the insignificant coefficient on Prize associates a prize increase from \$0.1 to \$3.8 with a human-score increase of 1.15. Of course, while we cannot reject a zero effect of the monetary prize, by the same token, and with the same p-value (p = 0.073), we also cannot reject an effect that is twice our point estimate, namely, 0.620—which would associate a prize increase from \$0.1 to \$3.8 with a human-score increase of 2.30. Note that for the EBRD model considered in GP and in Section IV above to accommodate **H1b** without **H1a** and **H1c**, participants have to care about winning each round's lottery regardless of its prize amount. Formally, v would be replaced by a constant v^* that represents the value of winning the lottery, and is unaffected by the monetary prize. We return to this point in the Discussion (Section VIII).[5]

C. Effect of Robot Effort on Human Attitudes

We used the data gathered from the three single-item scales asked after every round, as described in Section VI-A, to measure the effect of the robot's performance on robot competence, robot likability and the human's self competence.

To test **H2a–c**, we ran three fixed-effects regressions with the same specification, namely with robot score and prize as predictors, controlled for participant ID and round number.

- Table IV shows the regression estimates for robot competence. Robot score positively predicted the robot's competence ($\beta = 0.039,\ p < 0.0001$). Figure 3a illustrates the variation in mean robot competence with robot score.
- Similarly, Table V shows that robot score negatively predicted the robot's likability, $\beta = -0.016,\ p < 0.0001$). Figure 3b illustrates the variation in robot likability.
- Finally, Table VI shows that robot score negatively predicted the human's self competence ($\beta = -0.023,\ p < 0.0001$). Figure 3c illustrates the relevant variation.

[5]GP found some support for all three hypotheses (see their Table 2 and its discussion), and therefore for the EBRD model they consider. Specifically, in either a fixed- or random-effects specification, they estimated a negative interaction coefficient ($\beta = -0.049$ and $-0.051, p = 0.030$ and 0.037, respectively). Reproducing our Table III from their publicly available data, we estimate an average discouragement effect of the First Mover's on the Second Mover's effort ($\beta = -0.045, p = 0.089$) and an average prize-amount effect ($\beta = 0.44, p = 0.004$). While the two experiments are not directly comparable (e.g., the tasks and other design details are different), we note that GP's average discouragement effect is almost identical to ours.

TABLE IV: Effect of Robot Score on Robot Competence

Variable	Estimate	Std. Error	t-Ratio	p-Value
Robot Score	0.039	0.002	17.91	< 0.0001
Prize	0.037	0.027	1.35	0.176

$(n = 599)$

TABLE V: Effect of Robot Score on Robot Likability

Variable	Estimate	Std. Error	t-Ratio	p-Value
Robot Score	-0.016	0.002	-7.74	< 0.0001
Prize	-0.023	0.026	-0.89	0.375

$(n = 599)$

D. Individual Differences

We analyzed the data obtained from questionnaires measuring participants' self-reported self-efficacy [32], money-belief [33] and competitiveness [34], to study whether individual differences directly affected the performance or subjective measures, or interacted with our main hypotheses.

We did not have specific a-priori directional hypotheses, but tested the collected data along the following questions: Would people who value money as more important be more affected by the reward? How does their money belief affect their performance overall? Do people who had a higher belief in their ability to succeed (self-efficacy) feel better about their performance? How does this relate to the robot's performance? Do more competitive people exert a higher effort when the robot's performance increases? The last one was interesting since GP tested effort in a sequential competition, whereas our experimental paradigm included a simultaneous competition. People's tendency toward competitiveness could counteract the economic discouragement effect.

To evaluate these questions, we ran the following fixed-effects multivariate regressions:[6]

- Human effort as a function of money belief, prize, and their interaction, controlling for robot score and round number. Table VII shows the regression estimates. Money belief positively (suggestively) predicted the human's score ($\beta = 0.652,\ p = 0.007$), while the interaction between money belief and prize did not.

[6]In all of these regressions, the variables in the interaction terms were centered around their means.

TABLE VI: Effect of Robot Score on Self Competence

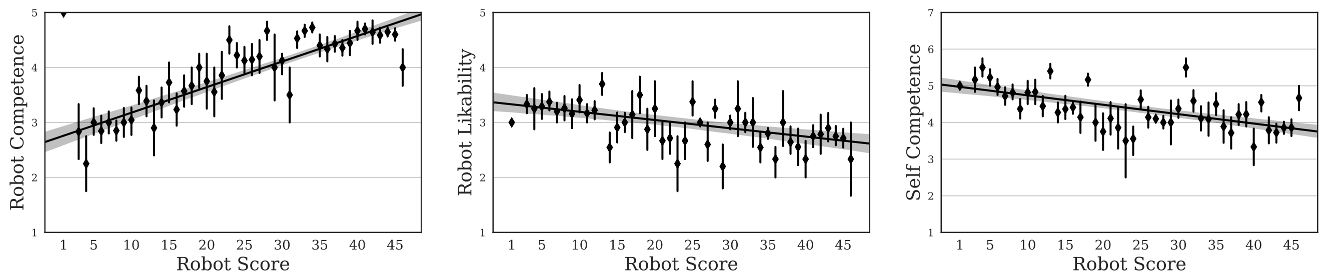
Variable	Estimate	Std. Error	t-Ratio	p-Value
Robot Score	-0.023	0.002	-10.05	< 0.0001
Prize	-0.038	0.028	-1.33	0.1846

$(n = 599)$

TABLE VII: Effect of Money Belief on Human Effort

Variable	Estimate	Std. Error	t-Ratio	p-Value
Money Belief	0.652	0.240	2.71	0.007
Prize	0.357	0.209	1.71	0.088
Money Belief $\times$ Prize	-0.135	0.205	-0.65	0.513

$(n = 600)$



(a) Robot Competence vs Robot Score (b) Robot Likability vs Robot Score (c) Self Competence vs Robot Score

Fig. 3: Effect of the robot's performance on human attitudes. (Each error bar is constructed using 1 standard error from the mean. The confidence interval for the fitted line is 99.5%. $n = 599$.)

- Self competence as a function of self-efficacy, robot score, and their interaction, controlling for prize and round number. As Table VIII shows, self-efficacy positively predicted the human's self competence ($\beta = 0.429, p = 0.002$), while the interaction between self-efficacy and robot score did not strongly predict it.

TABLE VIII: Effect of Self-efficacy on Self Competence

Variable	Estimate	Std. Error	t-Ratio	p-Value
Self-efficacy	0.429	0.136	3.16	0.002
Robot Score	-0.026	0.003	-7.65	< 0.0001
Self-eff. $\times$ Robot Sc.	0.018	0.009	1.92	0.055

($n = 599$)

- Human effort as a function of competitiveness, robot score, and their interaction, controlling for prize and round number. As Table IX shows, neither competitiveness nor its interaction with robot score predicted the human's score.

TABLE IX: Effect of Competitiveness on Human Effort

Variable	Estimate	Std. Error	t-Ratio	p-Value
Competitiveness	0.299	0.199	1.50	0.134
Robot Score	-0.027	0.016	-1.62	0.106
Compet. $\times$ Robot Sc.	-0.015	0.013	-1.14	0.254

($n = 600$)

E. Open-ended Responses

In their post-experiment responses, participants gave mixed feedback about their experience. Some said that they liked competing with the robot, while others said that it was stressful and frustrating:

> P014: *"It was an interesting task, I've never competed with a robot before. It was fun."*
>
> P048: *"I felt very stressed competing with the robot. In some rounds, I kept seeing the robot's score increasing out of the corner of my eye, which was extremely nerve-racking [sic]."*

Some reported that the prize and the robot's score affected how strongly they competed, while others reported that they chose to compete strongly in all the rounds, without regard for the speed of the robot or the prize. For most participants who reported that they were affected by the robot, they noted that they competed stronger when the robot was slower, as predicted by the theory:

> P035: *"[In] some rounds the robot would go slower and that's when I started going faster."*
>
> P055: *"Usually, if I saw a high expected point value, I worked less hard as my efforts would have less gain than a competition with a lower expected point value of the robot."*
>
> P011: *"I skipped rounds that were either too low in payout or too high in difficulty."*

Although in a minority of cases, the motivation was reversed:

> P031: *"When I had the lowest chance of winning, I tried my hardest to get it up."*

We found very little occurrences of anthropomorphizing the robot, such as:

> P064: *"It was obvious when the robot was 'going easy' on me."*

Most participants viewed the robot as a machine at best, or merely as a score to beat:

> P011: *"I actually did not really view this as competition with a robot, since the robot's predicted score was shown at the beginning of the round. I viewed it more as a challenge to myself to get as close to that number as possible."*
>
> P067: *"I sort of realized, I am just competing with an idea of mechanization, and the arm is just a prop to signify it. In reality, it's just a counter going up at a steady rate."*

Finally, a few of the participants complained about the robotic voice that read out the probability of winning every five seconds and noted that it was distracting.

F. Other Anecdotal Results

Further data analysis revealed some anecdotal results which could be grounds for future research. We ran a fixed-effects linear regression of the human's score as a function of self-efficacy, controlled for round number, prize, and robot score.

Self-efficacy negatively predicted the human's score, ($\beta = -2.396$, Std. Error $= 0.649$, t-Ratio $= -3.69$, $p = 0.0002$).

We had also recorded the number of mistakes made by participants in each round. We ran a fixed-effects regression for incorrect attempts as a function of robot score and prize, controlled for participant ID and round number. The regression results are shown in Table X. We see that there is no statistically significant effect of the robot's score or the prize on the number of mistakes made by the participants.

TABLE X: Effect of Prize and Robot Effort on # Errors

Variable	Estimate	Std. Error	t-Ratio	p-Value
Robot Score	0.006	0.003	1.72	0.086
Prize	0.039	0.044	0.90	0.371

($n = 600$)

VIII. Discussion

From an economic-theory point of view, our failure to find meaningful support for **H1a**—the prediction of a positive effect of prize amount on human effort—is rather puzzling. To accommodate it within a rational theory of decision-making, people would have to be willing to exert effort competing with the robot mostly for non-monetary reasons. Ex-ante, we did not expect such behavior. Ex-post, we could think of several potential explanations unique to our setup.

First, our experimental design itself might have caused participants to pay more attention to the probabilities of winning the prize than to the monetary amounts. Recall that in our experiment, a robotic voice read out loud, every five seconds, the updated probability of winning. This feature did not exist in any previous experiment we are aware of. Its purpose was to make sure that participants kept paying attention, while competing, to their dynamically changing probability of winning. But this may have caused participants to pay less attention to the monetary prize. Additional support for this possibility was mentioned in Section VII: A few participants complained about this feature and noted that it was distracting.

Another related important factor, and the primary difference between previous real-effort experiments and our experiment was the nature of the competition. To the best of our knowledge, this was the first experiment in which humans competed against a robot for a monetary prize, and where both the competitors worked simultaneously on the same task. Seeing a robot compete side-by-side may have motivated participants to focus on the competition and the chance of winning, without bothering about the monetary rewards. These explanations could be tested in future work.

Beyond this economic puzzle, we found a discouragement effect (**H1b**), that is similar to the average effect found by GP in human-human competition. We also found interesting related effects of the robot's performance on the participants' attitudes towards the robot and towards themselves. First, participants did not like a faster competitor robot as compared to a slower one even though they found the faster robot to be more competent. This may support an intuition that people like a weaker competitor, even if it is a robot.

Second, in line with **H2c**, participants' perception about their own ability to do this task was also negatively affected by the robot's performance—even though there was no direct interaction between the human tasks and that of the robot's. People perceived themselves as more competent when the robot was slower and as less competent when the competitor robot was faster. This may suggest that people assess their ability to perform a task relative to that of their competitor, even if it is a robot. Remarkably, this effect is present across rounds within subjects, in spite of the task itself never changing, with the only difference between rounds being the robot speed. This effect is not mainly driven by their lower score resulting from the above discouragement effect: Including the human's score as an additional explanatory variable in Table VI does not change the estimated effect of robot score on self competence more than trivially (it changes from -0.023 to -0.021; standard error remains 0.002). In other words, people's self competence ratings were negatively associated with the robot's performance even when controlling for their own performance (which was itself affected by that of the robot). As a side-note, while participants' overall self-efficacy significantly predicted their self-competence rating ($\beta = 0.429, p = 0.002$), this did not interact with the robot's score.

Finally, we assumed throughout that the human's score in each round represented their actual effort on the task. However, in our experimental design, it was possible for the participants to commit mistakes, which might have affected their scores. We did not find any statistically significant relation between the robot's score or the prize and the incorrect attempts made by participants. Thus, we maintain the working assumption that the human's score in each round is a reasonable measure of their actual effort on the task.

IX. Conclusion

To the best of our knowledge, this is the first experiment in which people competed with a robot, both working simultaneously on repetitive tasks for a monetary reward. We observed a small discouragement effect of the robot's performance on the human's performance. But we did not find a statistically significant effect of monetary reward on the human's performance. The latter is not easy to accommodate within any generally useful economics model. That said, an EBRD model where individuals value winning lotteries rather than the monetary prizes they deliver, appears consistent with our results of a discouragement effect with no money effect. While unique features of our experimental setup, such as the built-in salience of probabilities versus prize amounts, may make such a model easier to digest, it is unlikely to be usefully portable to more than a handful of other situations.

We found that participants liked a low-performing competitor robot more than a high-performing one, even though they considered the latter to be more competent. Participants' perception of their ability to do well on this task was also affected by the robot's performance. They considered themselves less competent when the robot performed better, even when controlled for their own score.

REFERENCES

[1] G. Hoffman and C. Breazeal, "Collaboration in human-robot teams," in *AIAA 1st Intelligent Systems Technical Conference*, 2004, p. 6434.

[2] J. A. Shah, P. R. Conrad, and B. C. Williams, "Fast distributed multi-agent plan execution with dynamic task assignment and scheduling." in *ICAPS*, 2009.

[3] E. C. Grigore, O. Mangin, A. Roncone, and B. Scassellati, "Predicting supportive behaviors for human-robot collaboration," in *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*, 2018, pp. 2186–2188.

[4] J. Schaeffer, N. Burch, Y. Björnsson, A. Kishimoto, M. Müller, R. Lake, P. Lu, and S. Sutphen, "Checkers is solved," *Science*, vol. 317, no. 5844, pp. 1518–1522, 2007.

[5] H. J. Berliner, "Backgammon computer program beats world champion," *Artificial Intelligence*, vol. 14, no. 2, pp. 205–220, 1980.

[6] M. Campbell, A. J. Hoane Jr, and F.-h. Hsu, "Deep Blue," *Artificial Intelligence*, vol. 134, no. 1-2, pp. 57–83, 2002.

[7] D. Silver, J. Schrittwieser, K. Simonyan, I. Antonoglou, A. Huang, A. Guez, T. Hubert, L. Baker, M. Lai, A. Bolton *et al.*, "Mastering the game of Go without human knowledge," *Nature*, vol. 550, no. 7676, p. 354, 2017.

[8] J. Wainer, K. Dautenhahn, B. Robins, and F. Amirabdollahian, "Collaborating with Kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism," in *10th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2010, pp. 631–638.

[9] S. Shahid, E. Krahmer, M. Swerts, and O. Mubin, "Child-robot interaction during collaborative game play: Effects of age and gender on emotion and experience," in *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. ACM, 2010, pp. 332–335.

[10] B. Mutlu, S. Osman, J. Forlizzi, J. Hodgins, and S. Kiesler, "Perceptions of ASIMO: an exploration on co-operation and competition with humans and humanoid robots," in *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction (HRI)*, 2006, pp. 351–352.

[11] E. Short, J. Hart, M. Vu, and B. Scassellati, "No fair!! An interaction with a cheating robot," in *5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 2010, pp. 219–226.

[12] T. O'Donoghue and J. Somerville, "Modeling risk aversion in economics," *Journal of Economic Perspectives*, vol. 32, no. 2, pp. 91–114, 2018.

[13] O. Morgenstern and J. Von Neumann, *Theory of games and economic behavior*. Princeton University Press, 1953.

[14] M. Rabin, "Risk aversion and expected-utility theory: A calibration theorem," in *Handbook of the Fundamentals of Financial Decision Making: Part I*. World Scientific, 2013, pp. 241–252.

[15] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the fundamentals of financial decision making: Part I*. World Scientific, 2013, pp. 99–127.

[16] D. E. Bell, "Disappointment in decision making under uncertainty," *Operations research*, vol. 33, no. 1, pp. 1–27, 1985.

[17] G. Loomes and R. Sugden, "Disappointment and dynamic consistency in choice under uncertainty," *The Review of Economic Studies*, vol. 53, no. 2, pp. 271–282, 1986.

[18] F. Gul, "A theory of disappointment aversion," *Econometrica: Journal of the Econometric Society*, pp. 667–686, 1991.

[19] B. Kőszegi and M. Rabin, "A model of reference-dependent preferences," *The Quarterly Journal of Economics*, vol. 121, no. 4, pp. 1133–1165, 2006.

[20] B. Kőszegi and M. Rabin, "Reference-dependent risk attitudes," *American Economic Review*, vol. 97, no. 4, pp. 1047–1073, 2007.

[21] B. Kőszegi and M. Rabin, "Reference-dependent consumption plans," *American Economic Review*, vol. 99, no. 3, pp. 909–36, 2009.

[22] D. Gill and V. Prowse, "A structural analysis of disappointment aversion in a real effort competition," *American Economic Review*, vol. 102, no. 1, pp. 469–503, 2012.

[23] T. ODonoghue and C. Sprenger, "Reference-dependent preferences," *Handbook of Behavioral Economics-Foundations and Applications 1*, p. 1, 2018.

[24] O. Heffetz, "Are reference points merely lagged beliefs over probabilities?" National Bureau of Economic Research, Tech. Rep., 2018.

[25] J. Bruce, T. Balch, and M. Veloso, "Fast and inexpensive color image segmentation for interactive robots," in *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 3. IEEE, 2000, pp. 2061–2066.

[26] S. Chitta, I. Sucan, and S. Cousins, "Moveit! [ROS topics]," *IEEE Robotics & Automation Magazine*, vol. 19, no. 1, pp. 18–19, 2012.

[27] P. Beeson and B. Ames, "TRAC-IK: An open-source library for improved solving of generic inverse kinematics," in *IEEE-RAS 15th International Conference on Humanoid Robots (Humanoids)*. IEEE, 2015, pp. 928–935.

[28] M. Quigley, K. Conley, B. Gerkey, J. Faust, T. Foote, J. Leibs, R. Wheeler, and A. Y. Ng, "ROS: an open-source Robot Operating System," in *IEEE International Conference on Robotics and Automation (ICRA) Workshop on Open Source Software*, vol. 3, no. 3.2. Kobe, Japan, 2009, p. 5.

[29] G. Hoffman and C. Breazeal, "Effects of anticipatory perceptual simulation on practiced human-robot tasks," *Autonomous Robots*, vol. 28, no. 4, pp. 403–423, 2010.

[30] C. Bartneck, D. Kulić, E. Croft, and S. Zoghbi, "Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots," *International Journal of Social Robotics*, vol. 1, no. 1, pp. 71–81, 2009.

[31] G. C. Williams and E. L. Deci, "Internalization of biopsychosocial values by medical students: a test of self-determination theory." *Journal of Personality and Social Psychology*, vol. 70, no. 4, p. 767, 1996.

[32] R. Schwarzer and M. Jerusalem, "The general self-efficacy scale (GSE)," *Anxiety, Stress, and Coping*, vol. 12, pp. 329–345, 2010.

[33] A. Furnham, "Many sides of the coin: The psychology of money usage," *Personality and Individual Differences*, vol. 5, no. 5, pp. 501–509, 1984.

[34] J. T. Spence, *Achievement and achievement motives : psychological and sociological approaches*. San Francisco : W.H. Freeman, 1983.

[35] C. G. Begley and J. P. Ioannidis, "Reproducibility in science: improving the standard for basic and preclinical research," *Circulation Research*, vol. 116, no. 1, pp. 116–126, 2015.

[36] D. J. Benjamin, J. O. Berger, M. Johannesson, B. A. Nosek, E.-J. Wagenmakers, R. Berk, K. A. Bollen, B. Brembs, L. Brown, C. Camerer *et al.*, "Redefine statistical significance," *Nature Human Behaviour*, vol. 2, no. 1, p. 6, 2018.