Abstract
The paper presents the results of a behavioral experiment conducted between February 2020 and March 2021 at Università Cattolica del Sacro Cuore, Milan Campus in which students were matched to either a human or a humanoid robotic partner to play an iterated Prisoner’s Dilemma. The results of a Logit estimation procedure show that subjects are more likely to cooperate with human rather robotic partners; that are more likely to cooperate after receiving a dialogic verbal reaction following the realization of a sub-optimal social outcome; that the effect of the verbal reaction is independent on the nature of the partner. Our findings provide new evidence on the effect of verbal communication in strategic frameworks. Results are robust to the exclusion of students of Economics related subjects, to the inclusion of a set of psychological and behavioral controls, to the way subjects perceive robots’ behavior and to potential gender biases in human-human interactions.

Keywords: Prisoner’s Dilemma, Communication, Human-Robot Interaction

JEL Codes: C91, D91

Acknowledgments
We acknowledge funding by two competitive research funds of Università Cattolica del Sacro Cuore: D3.2/2018 “Human-Robot confluence” and D3.2/2020 “Behavioral change”. The project is a spin-off of a previous project sponsored by Fetzer Institute “#3534.00” on the use of behavioral economics techniques for assessing behavioral change. We thanks M. Colagrossi, P. Natale, L. Stella, F. Trombetta for useful comments and observations on previous versions of the paper. The usual caveats apply. Research assistantship by E. Cerolini, P. Gambacciani, C. Marconi, F. Manzi and P. Zaza is acknowledged. Special mention is also due to F. Manzi whose expertise and skills were precious for programming the robot in the lab experiment.

Research assistantship by E. Cerolini, P. Gambacciani, C. Marconi, F. Manzi and P. Zaza is acknowledged. Special mention is also due to F. Manzi whose expertise and skills were precious for programming the robot in the lab experiment.
Track changes is off Everyone Track changes for everyone You Track changes for You mario.maggioni Track changes for mario.maggioni Guests Track changes for guests Current file Overview 26
1 Introduction

For decades human-robot interactions have been playing an important part in the production of cultural products (being stories, novels, movies or, more recently, videogames), populating the imaginary of people across countries and cultures. More recently, we are witnessing the spread of anthropomorphic robots outside movie sets or research labs, paving the way for an increasing role in economic and social interactions. These robots move and, most importantly, talk like human beings; this last feature is particularly relevant, since the ability to communicate through words makes a distinctive features of human interactions.

A long tradition in Philosophy since Socrates (469–399 BC) has stressed the ontological foundations of dialogic relationships. Martin Buber (1958) claimed that the full understanding of one own identity is strictly dependent on the dialogue with another presence.

More recently, Dumouchel and Damiano (2017) and Damiano and Dumouchel (2018) argued that dialogue is the fundamental structure and basic pattern of humans’ actions and thinking. Paleoanthropologists such as Dunbar and Bickerton (1996), Lieberman (2006) and Tattersall (2008) as well as paleoeconomists as Horan et al. (2005, 2008) formulated the hypothesis that verbal interactions and “verbal grooming” have been the crucial ingredients in the emergence of the evolutionary advantage of Homo Sapiens versus Homo Neanderthalensis, (thus explaining the survival of the former) some 150 thousands years ago; thus dialog has been somehow embedded in the evolutionary process of the human nature.\(^1\)

Further, communication has been empirically proved to increase trust and cooperation (Balliet, 2010), which in turn are pivotal in fostering positive social interactions. In particular, face-to-face communication has been shown to both promote and sustain cooperation across subjects even in strategic settings such as social dilemmas (Ostrom and Walker, 1997; Ostrom, 2000; Bicchieri, 2002).

In recent decades, the dialogic framework has been indeed enlarged in scope by the diffusion of anthropomorphic robots and the development of the sub-fields of Social Robotics (SR) and Human-Robot Interactions (HRI), which study mechanical objects able to communicate, verbally and non verbally, in similar ways to human beings and to play the

---

\(^1\)For a recent, at the moment isolated, dissenting voice, see Conde-Valverde et al. (2021).
role of artificial subjects, acting as “social partners” (Fong et al., 2003; Sung et al., 2010; Dumouchel and Damiano, 2017). Recent empirical research in HRI has implemented experimental frameworks in which human subjects have been partnered to humanoid robots in social dilemmas (e. g. Krach et al., 2008; de Melo et al., 2011; DeSteno et al., 2012; Paeng et al., 2016; de Melo and Terada, 2020). This process almost naturally led to ask whether the same may occur when people interacts with robotic agents, namely whether communication promotes cooperation also in such new settings.

To address this important issue in social science, in this paper, we devise a randomized experiment in which human subjects are randomly matched to either a human or an anthropomorphic robot partner and asked to perform an iterated Prisoner’s Dilemma (PD). In each of those two sub-samples, after the first round of the game, before being proposed to play the second round, half of the subjects are randomly assigned to treatment, which consists of a Dialogic Verbal Reaction (henceforth, DVR) performed by the partner, conditional on the occurrence of a social sub-optimum outcome in the first round of the PD.

The aim of our experiment is threefold: first, we investigate whether subjects behave differently depending on the nature of their partner (either human or robot); second, we analyze whether a DVR, implicitly referring to cooperation as a socially desirable strategy, influences the following choice of the subject; third, and most importantly, we check whether the effect of this verbal reaction depends on the (human or robotic) nature of the partner.²

Our results show that being exposed to a DVR, following sub-optimal social outcomes, positively affects subjects’ cooperation rate in the next stage and that this effect is independent on the partner’s type. In other words, despite the fact that all DVRs do not change the players’ payoffs, as long as the partner softly evokes cooperation - either in the form of an apology or reprimand or disappointment - the subject’s subsequent choice is bent towards cooperation, independently of it being a fellow human or an anthropomorphic robot.

This paper contributes to the existing economic literature in two ways. Firstly, we

²In a sense, our third research question can be thought as a sort of modified Turing Test. In the original three-person “imitation game” (Turing, 1950) an interrogator chats with two respondents, located in separate rooms, asking questions to detect which one of the two is a machine (being aware that one of them is a machine). If the interrogator cannot reliably tell the machine from the human, the machine is said to have passed the test. In our experiment the subject knows that the robot is a machine but he/she is surprised by the DVR and reacts as if the robot was a human.
provide new evidence on the effectiveness of communication in affecting decision-making, extending this result to HRI; secondly, and most importantly, we show that the effect of a DVR is independent on the human vs artificial nature of the agent, thus contributing to a still under-researched area in the study of economic interactions between human and non-human subjects.

2 Related literature

Several empirical works (e.g. Dawes et al., 1977; Braver and Wilson, 1986; Ostrom and Walker, 1991; Bicchieri, 2002) tested the relevance of communication in influencing the outcome of strategic interactions. Sally (1995) published the first meta-analysis of this stream of the literature and concluded that communication exerts the strongest effect, relative to other variables known to influence cooperation, such as group size, magnitude of the reward for defection, and group identity. Subsequently, Balliet (2010) addressed the same issue through a new meta-analysis and, by adopting some mediation-analysis techniques, confirmed that communication has a strong positive effect on cooperation within social dilemmas.

In a different epistemological tradition, “critical social theory”, Habermas relies on the notion of “strategic competence” in his account of the development of moral consciousness, defined as “the ability to make use of interactive competence for consciously processing morally relevant conflicts of action” (Habermas, 1979, p.88). He also refers to the ability of social actors discursively to examine the validity and legitimacy of established norms. Habermas thus “identifies a mechanism that might compellingly account for the binding force of language in strategic interaction” (Johnson, 1993, p. 81).

A number of recent studies investigated the potential effects of robots on human behavior. For example, on the one hand there is increasing concern that the complex relationship between human and machines may have detrimental effects on the mental health of workers (Robelski and Wischniewski, 2018) and might work as an additional stressor in the workplace (Körner et al., 2019). On the other hand, in some cases interactions with robots have shown to be perceived as skill enhancing and capable of increasing job satisfaction (Compagni et al., 2015) and that the individual and social perception of robots is influenced by physical appearance and behavior (Horstmann et al., 2018;
Marchetti et al., 2018; Manzi et al., 2021).

The nature of inter-subjects interactions has lately been extended to different types of agents by the diffusion of robots and, specifically, of humanoid social robots. The goal of robotics was to create “artificial workers” engaged in a broad range of activities.\(^3\) However, computer scientists and engineers soon had to acknowledge that, to perform in many of these fields, robots need to exhibit many social behaviors and, in particular, to evince a believable “social presence”, defined as a robot’s capability to give the user the “sense of being with another” (Biocca et al., 2003), or the “feeling of being in the company of someone” (Heerink et al., 2008). For this reason, social robots are able to express and/or perceive emotions; communicate with high-level dialogue; learn/recognize models of other agents; establish/maintain social relationships; use natural cues (gaze, gestures, etc.); exhibit distinctive personality and character; learn/develop social competencies.\(^4\)

The diffusion of humanoid robots and the emergence of the scientific sub-fields of “social robotics” and “human robot interactions” has spurred the interest in analyzing the possible consequences of repeated interactions of experimental subjects with “social robots” i.e. artificial agents that exhibit one or more of the previously defined “social” characteristics (Fong et al., 2003; Dumouchel and Damiano, 2017; Gaggioli et al., 2021).

Within the economic literature, the decision to trust another agent, to behave as a trustworthy partner in a transaction or to cooperate in a social dilemma\(^5\) is generally seen as not consistent with the pursuit of individual self-interest. For this reason, those decisions are usually explained either in terms of repeated interactions within a finite uncertain time horizon and appropriate time discount rate - for agents endowed with “self-regarding preferences” -, a certain degree of uncertainty over the type of the opponent, or by referring to concepts like equity and fairness for agents endowed with “other-regarding preferences”.

Relatively few papers explicitly consider the real relational dimension of agents’ inter-

---

\(^3\)&nbsp;Such as, among others, production, information, education, coaching, therapeutic mediation, assistance, entertainment, and companionship.

\(^4\)&nbsp;Indeed, a number of recent studies investigated the potential effects of robots on human behavior. For example, on the one hand there is increasing concern that the complex relationship between human and machines may have detrimental effects on the mental health of workers (Robelski and Wischniewski, 2018) and might work as an additional stressor in the workplace (Körner et al., 2019). On the other hand, in some cases interactions with robots have shown to be perceived as skill enhancing and capable of increasing job satisfaction. (Compagni et al., 2015)

\(^5\)&nbsp;Commonly operationalized within a game-theoretical framework through a Prisoner’s Dilemma, a Trust Game, a Centipede Game, or a “lost letter” experiment. See, among others, Dasgupta (1988), Kreps (1990) Yezer et al. (1996) and Skeath (1999).
teractions in social dilemmas. In those papers, cooperative and trustful behaviors are explained as motivated by the acknowledgment of the other party’s attitudes and intentions. Similarly, the literature on the so-called psychological games addresses the role of subjects’ intentions, by making payoffs belief-dependent (Geanakoplos et al., 1989; Rabin, 1993; Dufwenberg and Kirchsteiger, 2004; DeAngelo and McCannon, 2020).

Finally, support for the importance of attitudes, intentions and verbal and non-verbal cues in communication emerges both from behavioral economics, psychology and neuroeconomics experiments, where players show different behaviors and neurological activations when playing incentivized tasks and games with human counterparts as opposed to automata (from PCs to robots with various degrees of humanization), despite facing identical material payoffs (Kiesler et al., 1996; McCabe et al., 2001; Rilling et al., 2002, 2004; Bicchieri and Lev-On, 2007; Krach et al., 2008; Miwa and Terai, 2012; de Melo et al., 2011; Nouri and Traum, 2013; Paeng et al., 2016; Wu et al., 2016; Terada and Takeuchi, 2017; Crandall et al., 2018; Ishowo-Oloko et al., 2019).

3 Research design

Our research design addresses three specific questions that relates to the way humans interact with humanoid robots within a repeated Prisoner’s Dilemma.

**RQ1:** Do subjects’ cooperation rates differ across partner types? Following the recent developments in the empirical research in economics, psychology and social robotics we aim at understanding whether subjects facing a robotic rather than a human partner choose to cooperate with different or indistinguishable probabilities.

**RQ2:** Does a DVR affects the subjects subsequent behavior (independently of partner type) in terms of cooperation rates? Empirical evidence in the literature on social dilemmas shows that communication tends to promote cooperation. We aim at investigating whether subjects are more likely to cooperate after the partner has activated a DVR, after observing a sub-optimal outcome in the previous round of the game.

**RQ3:** Are subjects’ reactions to DVR depending on the partner’s nature (human vs robot)? This question directly originates from the former two. Since our
experiment is designed as a 2×2 matrix (see Table 1 below) we may expect that subjects’ decision are affected by either partner’s nature or DVR (or both of them). In other words, we want to explore whether any of these effects is characterized by heterogeneity.

To investigate how subjects behave when facing a robotic partner we need to exclude that their initial (online) choice in Phase 1 to cooperate or to defect is dependent on the partner’s type. At the same time, we cannot entirely rely on anonymized interactions, since the purpose of the investigation is strictly related with the possible effects arising from the interactions with different types of partner. For this reason, we designed an experimental procedure based on two distinct phases. In Phase 1, we asked the subjects (university students) to answer an online questionnaire and to play an incentivized task against an unknown anonymous partner. At the end of Phase 1 we asked subjects whether they wanted to come to the University Lab, proceeding to Phase 2, in order to: (i) know the result of the interaction; (ii) be rewarded; and (iii) possibly have other interactions with the partner.

We devised a 2×2 experimental design, as summarized in Table 1: on the one hand, we randomly assigned subjects to be matched with either a Human (H) or Robot (R) partner in the interactive situation (the prisoner’s Dilemma); on the other hand, we randomly administered a stimulus (treatment) to a fraction of our subjects both in the H and R experimental conditions.

The treatment consists in a “Dialogic Verbal Reaction” that the partner exhibits after observing a sub-optimal aggregate outcome of the interaction. Different stimuli were administered depending on the observed outcomes in the Prisoner’s Dilemma, as summarized in Table C2 in Appendix C.

---

Table 1: Experimental design

| Experimental condition | Treatment group |
|------------------------|-----------------|
| **Human**              | **No DVR**      |
| **Robot**              | **DVR**         |
| Baseline               | Reaction        |
| Robot                  | Interaction     |

---

6No DVR is activated when the aggregate Pareto optimal outcome is obtained, i.e. when both subject and partner cooperate.
3.1 Experimental procedures

The layout of our experiment is summarized by the flow-chart in Figure 1.

We sent invitation emails to all freshmen and sophomores students of *Università Cattolica del Sacro Cuore*, attending lectures in the Milan campus. The email included a link to an online survey containing the first phase of the experiment. Besides a set of psychological, attitudinal and socio-demographic questions, the survey included the first round of a repeated Prisoner’s Dilemma (as in Kreps et al., 1982). Each player is endowed with 3 euros and faces two strategies: if he/she chooses to “Cooperate”, his/her endowment is sent to the Partner who receive it doubled. If he/she chooses “Not Cooperate”, he/she keeps his/her endowment. Thus, the game proposed to subjects in this experiment consists of N repetitions of a two-person, two-strategy, round game (as summarized in Table 2).

Players, in each round, choose simultaneously. At the end of each round, players are informed of their own and their opponents’ choices, and shown the resulting payoffs. Total payoffs are the undiscounted sums of the round payoffs (plus the one-shot show-up fee).

---

7 *Università Cattolica del Sacro Cuore* is a multi-campus university mainly based in Milan but with other locations in Rome, Brescia, Piacenza and Cremona.

8 It seems reasonable to assume that IDR is not relevant for our experimental setting given that: (i) when subjects play their first round they are unaware about the delay between first and second round;
Table 2: Experiment’s payoff matrix

| Subject’s choice | Partner’s choice | Cooperate (C) | Not Cooperate (N) |
|------------------|------------------|---------------|-------------------|
| Cooperate (C)    | (6;6)            | (0;9)         |
| Not Cooperate (N)| (9;0)            | (3;3)         |

Students accepting to take part to the incentivized game were informed that they had been randomly paired to an unknown anonymous (human or artificial) partner who was playing the same game. If they accepted to participate, they were asked to make their choice (either Cooperate, C, or Not cooperate, N). The outcome of the experimental session (and the monetary reward, including show-up fee) would have been revealed only upon acceptance to show up and take part (in presence) to the second phase of the experiment at the university Lab. Only at that moment, respondents would discover whether the partner they faced in the online interaction was a Human or a Robot, the outcome of the interaction would be revealed and they might receive the invitation to play again one (or more) rounds of the game.⁹

Students who accepted to participate to Phase 2 had to schedule an appointment through an online third-party application. Phase 2, the proper Lab Experiment, consisted of three consequential stages. In Stage 1, the subject was introduced by the experimenter in the lab room where he/she could meet and face his/her partner, either Human or Robot¹⁰ with whom he/she was told he/she played the PD online in Phase 1. When the subject entered the experiment room, the robot was already there, while the human partner entered the room from the control room.

After having instructed the subject about the experimental procedures, the experimenter left the room. The subject would sit at a table, then the partner (either the Ph.D. student or the humanoid robot) greeted the subject and verbally introduced him/her/itself.¹¹

---

⁹The probability of being proposed a second round is 95%; the probability of being proposed a third round is 15%.

¹⁰The Robot was a NAO humanoid robot produced by Softbank Robotics (see Figure D4 in Appendix D for a picture, Gelin (2018) and Robaczewski et al. (2020) for references); human partners were Ph.D. students recruited and trained for this task.

¹¹In all sessions, Ph.D. students introduced themselves with their first name, greeted the subjects and asked for the subject’s first name; the robot, after introducing itself told the following in all experimental sessions: “Hi, I am Tom, a humanoid robot developed by Softbank Robotics able to perform complex
The partner then would tell the subject to wait and listen to the game director’s instructions,\textsuperscript{12} available both as a recorded voice over and shown on the screen in front of each player. Finally, the screen would reveal the outcome of the Prisoner’s Dilemma played online in Phase 1.

Based on the game outcome as shown in Table C2 in Appendix C, a random algorithm determined whether to activate the DVR with a 50\% probability, triggering the appropriate verbal stimulus performed by the partner, either human or robot.

The subject was then asked whether he/she would like to continue to play another round of the game, according to a random algorithm with a 95\% of positive probability. If the subject accepted, a second round was then implemented, without any further assignment of treatment. We checked that the probability of both proposal and acceptance to proceed to Stage 2 is not conditional on the outcome observed in Stage 1.\textsuperscript{13}

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{control_room_diagram.png}
\caption{Experiment room set}
\end{figure}

\textsuperscript{12}“Let’s wait for the instructions by the Game Director.”

\textsuperscript{13}We analyzed the association between a categorical variable, identifying all the 4 possible outcomes observed in Stage 1, and (i) a binary variable, identifying whether Stage 2 has been proposed to the subject; (ii) a binary variable, identifying whether this proposal has been accepted. In both cases, we can exclude any non-random pattern. In case (i), a Chi-sq. test ($Chi = 4.723, df = 3, p = 0.193$) and an ANOVA ($F - stat1.58, p = 0.194$) allow to infer that the proposal to continue to Stage 2 is interactions with human beings. What’s your name?”
After the results of the second round were revealed, the subject might have asked to continue with a further round of the game, according to a random algorithm with a 15% probability. If the subject wanted to play again, a third round of a PD was implemented, again without any treatment assigned.\(^\text{14}\)

After the last incentivized interaction, the game director communicated the total amount gained and asked to the subject whether he/she would like to donate 1 euro to a Charity of his/her choice.\(^\text{15}\)

Once this last choice was made, the subject had his/her receipt automatically printed, he/she left the lab room to enter the check-out area where he/she was paid the amount in cash. On leaving the Lab the student signed a confidentiality agreement requiring him/her not to spread any information about the experiment to other fellow students.

We planned to run the experiment from February, 13\(^{\text{th}}\) to March, 13\(^{\text{th}}\) 2020. However, due to the breakout of the COVID-19 pandemic, we had to stop all activities in the lab on February, 21\(^{\text{st}}\), that was our last day of data collection in 2020.

In February 2021, universities in Italy were open again for lectures in presence. We sized the moment\(^\text{16}\) and run another wave of the entire experiment (Phase 1 and Phase 2) from February 22\(^{\text{nd}}\) to March 31\(^{\text{th}}\). Again, due to the increase in the diffusion of the pandemic, we had to stop all activities in the lab on March the 4\(^{\text{th}}\).

\section{Data and estimation methods}

\subsection{The sample}

We sent a total of 23,552 emails, and received 2,205 unique answers and 1,893 valid and complete questionnaires (Phase 1). 1,829 respondents (96.6\% of valid observations) accepted to participate to the Second Phase of the experiment.

\(^\text{14}\)Stage 3 of the experiment was of no interest for the analysis. However, we devised the possibility to play 3 rounds of the PD to generate uncertainty over the total number of rounds in case of information spillover across students.

\(^\text{15}\)The available alternatives were \textit{Médecins Sans Frontières} and \textit{Greenpeace}.

\(^\text{16}\)Invitations were sent to freshmen and, to avoid duplication to those sophomores who did not open the invitation e-mail we sent them when they were freshmen in 2020.
The sample eligible for the analysis of treatment effect consists of 305 subjects making their choice in Round 2 after being assigned either to Treatment (Dialogi Verbal Reaction, DVR) or Control (No DVR) group (see table C4, top panel) out of a total of 490 subjects taking part to Phase 2 (bottom panel).\textsuperscript{17} A first inspection shows that subjects taking part to the research were overall more likely to Cooperate than to make a random choice: both in Stage 1 and Stage 2 the probability of choosing Cooperate is statistically different from 0.5 at a 0.1\% significance level.\textsuperscript{18}

Table 3 shows the summary statistics of both outcome and control variables by treatment group and experimental condition: the Robot and Human subsamples are shown in the top and bottom panel respectively; treatment and control values are shown across columns. As the table shows, all subsamples are well balanced across treatment groups in terms of control variables. Some imbalance occurs only for Female, less represented in the NO DVR group in the Robot condition (top panel) while more represented in the same group in the Human condition. A summary of balance tests, estimated through a Logit model, is shown in Appendix B in Figure B1, while Figure B2 reports the coefficients of balance tests also for partner types.\textsuperscript{19}

Interestingly, also the baseline measure of our outcome variable (i.e. the choice to Cooperate in Stage 1) is not statistically different across treatment groups. Conversely, a t-test shows that in both samples subjects assigned to the treatment group are more likely to cooperate than those in the control group, suggesting a potential effect of the DVR on subjects’ choices in Stage 2.

Finally, the bottom panel of Table 3 provides a summary of the differences in DVR effects by subsamples, i.e. it shows whether the increase in cooperation in the treatment group is heterogeneous across Human/Robot condition. Interestingly, being assigned to the Human condition is likely to increase the probability of cooperation in Stage 2 in the control group, where no DVR is performed; however, this statistical difference disappears when the comparison is made in the treatment group, suggesting that the effect of DVR

\textsuperscript{17}As illustrated in Table C2 in Appendix C a DVR stimulus could applied only in 3 out of 4 possible outcomes.
\textsuperscript{18}Please note that most of cooperative choices, i.e. those yielding an outcome of “CC” in Stage 1, were not included in Stage 2 in the Experimental sample: therefore, by design, the Experimental sample turns out to be “less cooperative” than the Full sample.
\textsuperscript{19}In this case, Freshman is statistically significant, and larger for Robot than Human partner. In Figure B3 we also provide a balance test for the two experimental waves, 2020 and 2021, showing that only Freshmen are more represented in 2021.
is not heterogeneous across partner types.

### Table 3: Summary statistics, by experimental conditions

|                  | DVR group | No DVR (control group) | T-test |
|------------------|-----------|------------------------|--------|
|                  | Mean      | St. Dev. | Obs | Mean | St. Dev. | Obs | Diff. | t-stat |
| **Outcome variables** |           |          |     |      |          |     |       |       |
| Cooperate (Stage 1) | 0.736     | 0.443    | 91  | 0.650| 0.480    | 80  | -0.086| (-1.22)|
| Cooperate (Stage 2) | 0.630     | 0.486    | 81  | 0.380| 0.489    | 71  | -0.249**| (-3.15)|
| **Control variables** |           |          |     |      |          |     |       |       |
| Female | 0.769 | 0.424 | 91  | 0.575| 0.497    | 80  | -0.194**| (-2.73)|
| Freshman | 0.857 | 0.352 | 91  | 0.800| 0.403    | 80  | -0.057 (-0.98)|
| Economics | 0.440 | 0.499 | 91  | 0.450| 0.501    | 80  | 0.010 (0.14)|
| Test failed | 0.099 | 0.300 | 91  | 0.075| 0.265    | 80  | -0.024 (-0.55)|

|                  | DVR group | No DVR (control group) | T-test |
|------------------|-----------|------------------------|--------|
|                  | Mean      | St. Dev. | Obs | Mean | St. Dev. | Obs | Diff. | t-stat |
| **Outcome variables** |           |          |     |      |          |     |       |       |
| Cooperate (Stage 1) | 0.640     | 0.483    | 86  | 0.663| 0.476    | 86  | 0.023 (0.32)|
| Cooperate (Stage 2) | 0.737     | 0.443    | 76  | 0.558| 0.500    | 77  | -0.178* (-2.34)|
| **Control variables** |           |          |     |      |          |     |       |       |
| Female | 0.628 | 0.486 | 86  | 0.802| 0.401    | 86  | 0.174* (2.57)|
| Freshman | 0.663 | 0.476 | 86  | 0.791| 0.409    | 86  | 0.128 (1.89)|
| Economics | 0.302 | 0.462 | 86  | 0.395| 0.492    | 86  | 0.093 (1.28)|
| Test failed | 0.151 | 0.360 | 86  | 0.186| 0.391    | 86  | 0.035 (0.61)|

**Robot vs. Human comparison**

|                  | Obs | Diff. | t-stat |
|------------------|-----|-------|--------|
| Cooperate (Stage 2): DVR | 157 | -0.11 | (1.44) |
| Cooperate (Stage 2): No DVR (control group) | 148 | -0.18** | (2.19) |

*Notes:* Summary statistics refer to subjects eligible for treatment, i.e. excluding those observing “CC” as Stage 1’s outcome.

### 4.2 Estimation technique

We address the aforementioned set of research questions by implementing a Logit model, in which the probability that respondent $i$ makes a Cooperative choice is conditional on a set of control variables and experimental conditions. Formally:

$$Y_{i,s} = \begin{cases} 
1 & \text{if respondent } i \text{ chooses Cooperation at round } r \\
0 & \text{if respondent } i \text{ chooses Non-Cooperation at round } r 
\end{cases}$$

$$\log \left( \frac{\pi_{i,s}}{1 - \pi_{i,s}} \right) = \beta_0 + \delta_1 Robot_i + \delta_2 DVR_i + \delta_3 Robot_i \times DVR_i$$

$$+ \beta_4 Choice_{i,s-1} + \beta_j X_{j,i} \cdots + \beta_k X_{k,i} \quad \text{(1)}$$
where \( \pi_i \) is the probability that \( Y_i \) equals 1; \( Y \) is the dependent variable measured at rounds 1 and 2; \( \beta_0, \beta_1, \beta_j, \ldots, \beta_k \) and \( \delta \) are the parameters to be estimated, with \( \delta_1, \delta_2 \) and \( \delta_3 \) being respectively the main coefficients of interest for RQ1, RQ2 and RQ3; \( \text{Choice}_{i,s-1} \) is the choice made by respondent \( i \) in the previous experimental stage, \( s-1 \) and \( X_j \cdots X_k \) are a set of \( k \) experiment-related and control variables, illustrated below.

To increase the precision of the estimates and to account for potential confounding factors, all models are estimated through different specifications. Control variables include Experiment-related controls and background characteristics of the subjects. Among the former set of controls we include \textit{Instruction order}, a categorical variable that takes into account which of the potential PD’s outcomes is shown first to the subject during the introductory instructions of the game, to control for potential priming;\(^{20} \text{Wave} \), to control for the year of the experiment (either 2020 or 2021); \textit{Experiment day} to account for the number of days the experiment has been running (e.g. Day 1, 2, 3, ...) to control for potential spillover effects. Among the latter group of controls we include \textit{Female}, to control for subjects’ gender; \textit{Freshman}, to control for subjects attending first year of their BA; \textit{Econ}, to control for subjects enrolled in BAs in Economics, Management, Finance and Banking; \textit{Fail test}, to control for subjects failing the pre-game test designed to verify their full understanding of the game’s procedures, rules and payoffs. Moreover, all models are also estimated by excluding this latter group of subjects. All variables’ description is summarized in Table C3 in Appendix C.

Finally, since our experiment design consists of \( 2 \times 2 \) treatment conditions, in Section 6 we test the first two hypotheses by estimating separately the effects of \textit{Robot} and \textit{DVR} by resorting to sample splits.

5 Results

5.1 Cooperation patterns

On average, all subjects eligible for treatment who proceed to Stage 2 choose to Cooperate 58% of the times, as shown in the top panel in Table C4. The breakout by treatment and

\(^{20}\text{Subjects received the illustration of all potential outcomes of PS in random order. We control for the outcome that appears first to the subject, to account for potential priming effect of instructions. We also replicated our results recoding this variable to account for the outcome that has been shown last. Results do not change.} \)
experimental condition is illustrated in Figure 3 and summarized in Table C1 in Appendix C.

As Table C1 shows, in the Control group, where no DVR is assigned to either Human or Robot partner, the average cooperation rate is lower in the Robot condition.

On the other hand, in the Treatment group subjects facing either a Human or a Robotic partner which implement a DVR, display higher cooperate rates. Overall, this descriptive patterns suggest that in the absence of any verbal reaction after the observed outcome in Stage 1 subjects are more likely to cooperate when facing a Human Partner: however, if a verbal reaction (DVR) is performed by either a Robotic or Human partner, subjects are more likely to Cooperate than in the no-reaction case. These patterns are visually summarized by Figure 3, in which the bars’ height is clearly higher in the DVR group for both Human and Robot conditions.

Our main hypotheses are tested through the estimation of Eq. 1, whose outcome is shown in Table 4. In this table there are three main coefficients of interest: Robot, DVR and DVR×Robot.
5.2 The effect of Partner type

The estimated coefficient of Robot in Table 4 provides evidence showing that subjects display different behavior depending on partner’s types: being assigned to a Robot significantly decreases the cooperation rate, with an estimated probability between 16.5% and 18%, depending on specifications, as shown in the “Margins” columns. This result holds once control variables are included, as shown in Table C6 in Appendix C, and the magnitude of the coefficient is substantially the same, implying a lower probability to cooperate between 15.6% and 18.3%. Overall, these results show that, ceteris paribus, cooperation is strongly and significantly affected by the partner type. This result is aligned with the evidence provided by a systematic review of computer players performed by March (2019, p. 3) who finds that: (i) subjects’ “behavior often changes when human opponents are replaced by computer players; (ii) subjects generically behave more selfish and more rational when interacting with computers; (iii) subjects often learn to exploit computer players, even if the latter do not follow a fixed strategy but are responsive to the subjects’ choices, and if subjects possess little information about computers a priori.”

Table 4: Main results: the effect of DVR and Robot, experimental sample

| Choice: “Cooperate” | Benchmark Including Choice at Stage 1 |
|---------------------|--------------------------------------|
|                     | Full sample Excl. failed tests† Margins | Full sample Excl. failed tests† Margins |
| DVR × Robot         | 0.117 (0.503) 0.297 (0.545) 0.028 (0.122) | -0.017 (0.525) 0.233 (0.569) -0.004 (0.127) |
| DVR                 | 0.843 (0.363)** 0.776 (0.401)* 0.204 (0.088)** | 0.926 (0.378)** 0.793 (0.419)* 0.223 (0.091)** |
| Robot               | -0.681 (0.355)* -0.874 (0.385)** -0.165 (0.086)* | -0.762 (0.370)** -1.049 (0.406)** -0.184 (0.089)** |
| Cooperate (Stage 1) | 1.331 (0.288)** 1.427 (0.312)** 0.321 (0.069)** |

Inst. order | Wave | Exp. day | Pseudo R-sq. | Obs | LL | AIC | BIC
| Yes | Yes | Yes | 0.07 | 305 | -192 | 416 | 476 |
| Yes | Yes | Yes | 0.09 | 268 | -167 | 366 | 424 |
| Yes | Yes | Yes | 0.13 | 305 | -181 | 395 | 459 |
| Yes | Yes | Yes | 0.15 | 268 | -156 | 346 | 407 |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Give-Some-Game. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses; bootstrap standard errors between brackets.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* p < 0.10, ** p < 0.05, *** p < 0.01

Electronic copy available at: https://ssrn.com/abstract=3834619
5.3 The effect of DVR (Dialogic Verbal Reaction)

The coefficient of DVR provides evidence supporting that DVR positively affects cooperation rates: in all models, subjects cooperate more when facing a DVR, with a probability between 20.4% and 22.3%. Also in this case, the results hold once control variables are included, as shown in Table C6 in Appendix C. All types of DVR exerts a positive influence on the cooperation rate, as summarized by Figure D3. However, the lack of statistical power, due to limited sub-sample size, hinders a disaggregate analysis. In aggregate, this result provide new evidence on the positive effect of communication in fostering cooperation in strategic decision-making.

5.4 Heterogeneity of DVR effect across partner types

In the previous section we have shown that the effect of DVR is strong and significant in both (Human and Robot) experimental conditions. In this section we test whether these effects are heterogeneous across partner types, i.e. whether facing a Human or Robot partner affects the way verbal interactions promote cooperative choices. In Table 4, the coefficient of the interaction variable $DVR \times Robot$ provides the outcome of this test.\(^{21}\) As the table shows, the coefficient is never significantly different from 0.

Note that the coefficients of both $Robot$ and $DVR$ are significantly different from 0, as shown in the previous sections. Therefore, while the effect of the DVR is large and significant, such effect is not heterogeneous across partner types. In other words, once the respondent is randomly assigned to the DVR treatment group, he/she is more likely to cooperate with his/her partner, irrespective of the fact that it is a Human or a Robot. The lack of heterogeneity provides a striking result: as far as the partner in the PD performs a DVR, subjects respond by increasing, on average, their probability to cooperate. Therefore, the ability to perform a verbal dialogic interaction - which indirectly evokes a commonly shared ethic norm in terms of the social desirability of cooperation - eliminates any previous differential cooperative attitudes with respect to human versus robots. These results are robust to the inclusion of the usual set of control variables, (as shown in Table C6, in Appendix C).

\(^{21}\)Ai and Norton (2003) show some concerns about the actual interpretation of interaction effects in nonlinear models, such as Logit. For this reason, we replicate all our main results through a Linear Probability Model, estimated through Ordinary Least Squares (OLS). These replications are shown in Appendix Appendix A. All results are confirmed.
The results shown in Table 4 may depend on the belief that a robot which is able to implement a sensible contextual dialogical interaction\textsuperscript{22} is “more human”, thus it deserves from the subject a behavior reserved for fellow human partners. Alternatively, the DVR may produce the effect by simply acting as a mere soft reminder of the social desirability of cooperation.\textsuperscript{23} Our experimental framework is unable to disentangle between these two competing and/or complementary explanations. However, it is worth noting that NAO, the robot we used in the experiment, despite being able to display appropriate gaze and body gesture cues to increase its appearance of “socialness”, is still far away to reproduce pitch, accent and expressiveness of a natural human voice.\textsuperscript{24} Further research is needed in order to disentangle these two possible explanation. One possible strategy is to have the DVR performed by the Game Director (the pre-record voice over) commenting the realization of social sub-optimal results. In this the “ethical reminder” would have been separated from the partner (and, in particular, from the robot).

6 Robustness checks

In this section we provide robustness checks for the main result shown in the paper: while being matched with a robot partner decreases, \textit{ceteris paribus}, cooperation rates and receiving a DVR by the partner (of any type) increases them, the effect of DVR is not heterogeneous across partner type, i.e. it does not depend on the partner being a robot or a human.

We address potential sources of bias in our results that may be driven by the presence of Econ students; the subject’s psychological and behavioral traits; the subject’s perception about robots’ behavior; the occurrence of gender biases. We address each one of these concerns in separate subsections and finally also checked for possible differential outcomes when using sample splits.\textsuperscript{25}

\textsuperscript{22}Note that we can exclude that the effect is due to the mere fact of NAO being able to talk since it greets any subject he is interacting with, whether or not, in a later moment, it will perform a DVR.

\textsuperscript{23}Similarly to the role played by the mentioning of the Ten Commandments as in (Mazar et al., 2008), or the honor code as in (McCabe and Trevino, 1993) in stimulating academic honesty.

\textsuperscript{24}NAO communicated with subject via a “Wizard of Oz system” controlled by the laptop PC located in the control room. All moves and speech items were written and coded in the system. The robot followed a pre-programmed protocol where the experimenter did not need to speak or type anything during the interaction but only press a button to start the interaction as in Laban et al. (2021).

\textsuperscript{25}All our main results hold also implementing a difference-in-difference estimation strategy. Results are available upon request.
6.1 Excluding Econ students

To assess the robustness of our results, we perform our econometric analysis on different sub-samples. First of all, we analyze the potential bias induced by students attending BAs in Economics, Management, Finance and Banking (henceforth: Econ students). Empirical evidence shows that students in this population are more self-interested than their non-Econ peers either because of a self-selection process (Frey and Meier, 2005; Carter and Irons, 1991) or because of an indoctrination effect (Frank et al., 1993; Haucap and Müller, 2014).26 We are not interested here in analyzing the causes of this differential behaviors; nevertheless, we tested whether our results are driven by the inclusion of a substantial number of Econ students which accounts for 36% of the sample.27 Table 5 displays the results of the estimation of Eq. (2), as in Section 5.4 when Econ students are excluded from the analysis. The main outcome of the analysis is confirmed, since the coefficient of the interacted variables is still not significant, both the size and significant of DVR and Robot coefficient is consistent with the results shown in Section 5.4. To complete our analysis, in Table C7 in Appendix C we show results of the same models when control variables are included. Also in this case, the main outcome of our analysis still hold.

6.2 Controlling for psychological and behavioral factors

As a further robustness check we assess whether our results are affected by underlying unobserved psychological and behavioral traits of subjects. The choice to cooperate entails the risk of being defected and it may be driven by the level of trust “embedded” in each individual subject. Further, more or less impulsiveness may drive the choice towards defection or cooperation. Finally, cooperation may be related to subjects’ level of pure altruism. For this reason, Table 6 shows the outcome of the same models estimated in Section 5.4 by controlling for a set of psychological scales based on subjects’ answers in Phase 1 (before the lab experimental session took place). These scales include a Risk Scale, retrieved from the General Risk Scale developed by Dohmen et al. (2011), a Trust Scale, retrieved from the Trust in People Scale (Inter-University Consortium for Political Research, 1964), a simple 12-items matching measure of Intertemporal Discount Rate.

26 A few studies (Yezer et al., 1996; Hu and Liu, 2003) using different incentivized situations, find evidence that Econ students are more willing to cooperate.
27 This percentage closely matches the proportion (37%) of Econ Students on the total student population of the Milan Campus of the Catholic University.
Table 5: Robustness check: excluding Econ students

| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|----------------------------|
|                     | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
| DVR x Robot         | 0.016      | 0.420                   | 0.004   | -0.084      | 0.453               | -0.019  |
|                     | (0.664)    | (0.732)                 | (0.154) | (0.690)    | (0.762)             | (0.157) |
| DVR                 | 1.134      | 0.975                   | 0.262   | 1.163       | 0.900               | 0.265   |
|                     | (0.472)**  | (0.528)*                | (0.108)**| (0.495)**  | (0.554)             | (0.111)**|
| Robot               | -0.648     | -0.956                  | -0.150  | -0.825      | -1.308              | -0.188  |
|                     | (0.475)    | (0.524)*                | (0.110) | (0.488)*    | (0.555)**            | (0.111)*|
| Cooperate (Stage 1) | 1.380      | 1.573                   | 0.315   |
|                     | (0.409)**  | (0.466)**               | (0.093)**|
| Inst. order         | Yes        | Yes                     | Yes     | Yes         | Yes                 | Yes     |
| Wave                | Yes        | Yes                     | Yes     | Yes         | Yes                 | Yes     |
| Exp. day            | Yes        | Yes                     | Yes     | Yes         | Yes                 | Yes     |
| Pseudo R-sq.        | 0.12       | 0.14                    |         | 0.17        | 0.20                |         |
| Obs                 | 192        | 166                     | 192     | 192         | 166                 | 192     |
| LL                  | -112       | -95                     |         | -106        | -89                 |         |
| AIC                 | 257        | 223                     |         | 246         | 212                 |         |
| BIC                 | 309        | 272                     |         | 302         | 265                 |         |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Students attending BAs in Economics, Management, Finance and Banking are excluded from the sample. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* p < 0.10, ** p < 0.05, *** p < 0.01

(IDR) based on Thaler (1981) and a “raw” measure of generosity (Donation) proxied by the decision to donate 1 euro out of their final reward to a Charity of their choice at the end of the experiment.

While the empirical literature has not reached a consensus on the relation between risk aversion and cooperation (Dolbear Jr and Lave, 1966; Sabater-Grande and Georgantzis, 2002; Burks et al., 2009; Proto et al., 2019), trust is generally associated with a positive influence on cooperation (Tedeschi et al., 1969; Cook and Cooper, 2003; Jung et al., 2012), and IDR and impulsiveness have been shown to be negatively correlated to cooperation (Yi et al., 2005; Streich and Levy, 2007; Jones and Rachlin, 2009; Locey et al., 2013; Malesza, 2020) As Table 6 shows, only the coefficient of Trust scale is weakly significant and consistent with the current empirical evidence. All the remaining Psychological and Behavioral controls are not statistically significant and the estimated coefficients relating to the main effects are unchanged with respect to the main model.
Table 6: Robustness check: controlling for psychological and behavioral factors

| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|-----------------------------|
|                     | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
| DVR × Robot         | 0.167      | 0.310                | 0.040   | 0.035      | 0.213               | 0.008   |
|                     | (0.500)    | (0.535)               | (0.121) | (0.520)    | (0.557)              | (0.126) |
| DVR                 | 0.836      | 0.787                | 0.292   | 0.928      | 0.854               | 0.244   |
|                     | (0.359)**  | (0.391)**            | (0.087)**| (0.375)**  | (0.409)**           | (0.090)**|
| Robot               | -0.726     | -0.834               | -0.176  | -0.798     | -0.981              | -0.193  |
|                     | (0.347)**  | (0.375)**            | (0.084)**| (0.359)**  | (0.392)**           | (0.087)**|
| Cooperate (Stage 1) | 1.226      | 1.296                | 0.296   | 1.226      | 1.296               | 0.296   |
|                     | (0.276)**  | (0.296)**            | (0.067)**| (0.276)**  | (0.296)**           | (0.067)**|

**Psychological and behavioral controls:**

|                      | Full sample | Excl. failed tests† | Margins |
|----------------------|-------------|---------------------|---------|
| Risk scale           | 0.009       | 0.010               | 0.002   |
|                      | (0.057)     | (0.060)             | (0.014) |
| Trust scale          | 0.235       | 0.214               | 0.057   |
|                      | (0.135)*    | (0.144)             | (0.033)*|
| IDR                  | -0.004      | -0.002              | -0.001  |
|                      | (0.066)     | (0.066)             | (0.001) |
| Donation             | 0.137       | 0.016               | 0.033   |
|                      | (0.250)     | (0.269)             | (0.061) |
| Inst. order          | Yes         | Yes                 | Yes     |
| Wave                 | Yes         | Yes                 | Yes     |
| Exp. day             | Yes         | Yes                 | Yes     |
| Pseudo R-sq.         | 0.07        | 0.08                | 0.12    |
| Obs                  | 305         | 268                 | 305     |
| LL                   | -192        | -169                | -182    |
| AIC                  | 408         | 362                 | 390     |
| BIC                  | 453         | 405                 | 438     |

**Notes.** Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

∗ p < 0.10, ** p < 0.05, *** p < 0.01

6.3 Controlling for perceptions about Robots

Facing a robotic partner in a PD is a currently an uncommon situation for human subjects. Since we do not know a priori how our subjects perceive the robot partner (especially in terms of preferences and/or beliefs), we decided to include a simple questionnaire during Phase 1 in which we asked subjects to identify their perception of robots. In this way we obtained information about whether subjects perceive the robot as an “Adaptive” device, or as a device simply running a predetermined set of program lines.28

Table 7 provides a test of the potential heterogeneity driven by the perception of the robot as an Adaptive tool. Interestingly, also in this case, the main outcome of

28The set of all possible answers is listed in Table C3. Also note that the vast majority of subjects answered either “Robots executes a fixed list of commands and operations” or “Robot adapts its behavior to the interaction with the human being”, thus supporting our binary coding.
our analysis is unchanged, with both DVR and Robot variable that remain statistically
significant, while all the interaction terms, capturing potential heterogeneity, are not
significantly different from 0. As a further check, we inspect whether the effect of DVR is
still significant in the Robot subsample, as shown in Table 8. The outcome shown in the
Table strengthens our main result, confirming that DVR is the main driver of cooperative
behavior in our experiment.

Table 7: Robustness check: controlling for heterogeneity driven by perception about
robot’s behavior

| Choice: “Cooperate”       | Benchmark | Including Choice at Stage 1 |
|---------------------------|-----------|----------------------------|
|                           | Full sample | Excl. failed tests | Margins | Full sample | Excl. failed tests | Margins |
| DVR×Robot×Adaptive        | -0.847     | -0.968                   | -0.205   | -1.271     | -1.260                   | -0.307   |
|                           | (1.014)    | (1.093)                  | (0.246)  | (1.063)    | (1.149)                  | (0.257)  |
| DVR×Robot                 | 0.489      | 0.740                    | 0.119    | 0.529      | 0.793                    | 0.128    |
|                           | (0.669)    | (0.722)                  | (0.162)  | (0.697)    | (0.752)                  | (0.168)  |
| DVR×Adaptive              | -0.104     | -0.083                   | -0.0251  | 0.115      | -0.011                   | 0.028    |
|                           | (0.727)    | (0.799)                  | (0.176)  | (0.762)    | (0.843)                  | (0.184)  |
| Adaptive×Robot            | 0.954      | 1.021                    | 0.231    | 0.947      | 0.844                    | 0.228    |
|                           | (0.711)    | (0.785)                  | (0.172)  | (0.737)    | (0.823)                  | (0.178)  |
| DVR                       | 0.885      | 0.805                    | 0.215    | 0.889      | 0.808                    | 0.215    |
|                           | (0.470)*   | (0.518)                  | (0.115)* | (0.493)*   | (0.536)                  | (0.119)* |
| Robot                     | -1.080     | -1.302                   | -0.262   | -1.152     | -1.406                   | -0.278   |
|                           | (0.468)**  | (0.511)**                | (0.114)**| (0.484)**  | (0.532)**                | (0.117)**|
| Adaptive                  | -0.251     | -0.314                   | -0.061   | -0.267     | -0.155                   | -0.065   |
|                           | (0.488)    | (0.556)                  | (0.118)  | (0.509)    | (0.591)                  | (0.123)  |
| Cooperate (Stage 1)       | 1.350      | 1.447                    | 0.326    | (0.292)**  | (0.318)**                | (0.070)**|
| Wave                      | Yes        | Yes                      | Yes      | Yes        | Yes                      | Yes      |
| Inst. order               | Yes        | Yes                      | Yes      | Yes        | Yes                      | Yes      |
| Exp. day                  | Yes        | Yes                      | Yes      | Yes        | Yes                      | Yes      |
| Pseudo R-sq.              | 0.08       | 0.09                     | 0.14     | 0.14       | 0.16                     |         |
| Obs                       | 305        | 268                      | 305      | 305        | 268                      | 305      |
| LL                        | -191       | -166                     | -179     | -155       |                          |         |
| AIC                       | 421        | 371                      | 400      | 351        |                          |         |
| BIC                       | 496        | 443                      | 479      | 426        |                          |         |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing
the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses.
† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.
* p < 0.10, ** p < 0.05, *** p < 0.01
Table 8: Robustness check: controlling for perception about robot’s behavior, Robot sample

| Choice: “Cooperate” | Benchmark Including Choice at Stage 1 | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
|---------------------|--------------------------------------|-------------|---------------------|--------|-------------|---------------------|--------|
| DVR×Adaptive        |                                      | -0.754      | -0.918              | -0.188 | -1.033      | -1.124              | -0.258 |
|                     |                                      | (0.745)     | (0.790)             | (0.186) | (0.793)     | (0.840)             | (0.198) |
| DVR                 |                                      | 1.407       | 1.568               | 0.352  | 1.554       | 1.682               | 0.388  |
|                     |                                      | (0.501)***  | (0.539)***          | (0.125)*** | (0.541)***  | (0.581)***          | (0.135)*** |
| Adaptive            |                                      | 0.684       | 0.705               | 0.171  | 0.698       | 0.698               | 0.174  |
|                     |                                      | (0.538)     | (0.571)             | (0.134) | (0.564)     | (0.597)             | (0.141) |
| Cooperate (Stage 1) |                                      |             |                     |        | 1.610       | 1.543               | 0.402  |
|                     |                                      |             |                     |        | (0.463)***  | (0.486)***          | (0.116)*** |
| Wave                | Yes                                  | Yes         | Yes                 | Yes    | Yes         | Yes                 | Yes    |
| Inst. order         | Yes                                  | Yes         | Yes                 | Yes    | Yes         | Yes                 | Yes    |
| Exp. day            | Yes                                  | Yes         | Yes                 | Yes    | Yes         | Yes                 | Yes    |
| Pseudo R-sq.        | 0.15                                 | 0.16        |                     |        | 0.21        | 0.22                |        |
| Obs                 | 152                                  | 140         | 152                 |        | 152         | 140                 | 152    |
| LL                  | -90                                  | -81         |                     |        | -83         | -76                 |        |
| AIC                 | 212                                  | 195         |                     |        | 200         | 185                 |        |
| BIC                 | 260                                  | 242         |                     |        | 252         | 235                 |        |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Subjects assigned to Human partner not included in the sample. Standard errors between parentheses. † This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

∗ p < 0.10, ** p < 0.05, *** p < 0.01

6.4 Controlling for gender biases

Since we engaged both male and female PhD students to perform the Human partner, a possible concern may arise from gender biased responses of subjects depending on both their own and their partner’s gender and by a combination of the two. For this reason, we run a robustness check by including both subject’s and partner’s gender as well as an interactive term of the two in a new model, whose outcomes are presented in Table 9. This analysis is clearly performed only on the Human subsample. As the results in the Table show, the main outcome of our analysis is confirmed: while DVR remains strongly significant, all interaction terms are not.

6.5 Sample splits

In this section we replicate a test of H1 and H2 by resorting to sample splits: since our experimental design involves a 2×2 treatment/condition matrix, we exploit sample splits to assess the robustness of our main results by testing H1 and H2 separately and inde-
Table 9: Test for potential gender bias in PD interactions, experimental sample

| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|----------------------------|
|                     | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
| DVR × Female × Female partner | -1.933 | -1.803 | 0.196 | 0.587 | 2.141 | -1.347 | -0.869 |
| DVR × Female | 0.057 | -0.162 | -1.179 | -1.549 | 0.0832 | -0.558 |
| DVR × Female partner | 1.289 | 0.745 | 0.0832 | -0.558 | (1.823) | (1.868) |
| Female × Female partner | 2.627 | 2.091 | 1.893 | 1.125 | 1.532 | 1.623 |
| DVR | 0.888 | 1.137 | 0.196 | 0.240 | 2.049 | 2.300 | (1.593) | (1.646) | (0.0860)*** |
| Female | -1.548 | -1.202 | -0.107 | -1.131 | -1.108 | -0.578 | (1.295) | (1.359) | (0.0951) |
| Female partner | -2.128 | -1.273 | -0.679 | -0.9728 | -1.607 | -0.651 | (1.381) | (1.461) | (0.121) |
| Cooperate (Stage 1) | 1.700 | 1.895 | 0.366 | (0.465)*** | (0.512)*** | (0.9898)*** |
| Inst. order | Yes | Yes | Yes | Yes | Yes | Yes |
| Wave | Yes | Yes | Yes | Yes | Yes | Yes |
| Exp. day | Yes | Yes | Yes | Yes | Yes | Yes |
| Pseudo R-sq. | 0.11 | 0.10 | 0.19 | 0.20 |
| Obs | 153 | 128 | 153 | 153 |
| LL | -88 | -75 | -86 | -67 |
| AIC | 216 | 189 | 204 | 175 |
| BIC | 277 | 246 | 267 | 235 |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses. Control variables include: Female, Freshman, Economics and Failed test.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.
* p < 0.10, ** p < 0.05, *** p < 0.01

Table 10 summarizes the results relating to the test of H1 (whether subjects behave differently depending on partner type) by treatment groups (i.e. in DVR and No DVR subsamples). The upper panel in Table 10 refers to the control group, where no DVR was performed by the partner. In this case, being assigned to a Robot significantly decreases the cooperation rate, with an estimated probability of about 18%, as shown in the “Margins” columns. A similar result can be found in the treated group, reported in the bottom panel, although in this case only when the choice made by the subject in Stage 1 is accounted for. In the Appendix, Table C5 shows that the result is robust to the inclusion of a set of control variables. Overall, the main results are confirmed: subjects tend to cooperate less when facing a robotic partner, especially when it does not perform a verbal reaction, thus suggesting that, when subjects face a robot (which does not verbally react to the game outcome), they are more likely to defect compared to their
behavior with a fellow human beings

Table 10: Effect of Partner type, by treatment condition, experimental sample

| No DVR condition | Benchmark | Including Choice at Stage 1 | Benchmark | Including Choice at Stage 1 |
|------------------|-----------|----------------------------|-----------|----------------------------|
|                  | Full sample | Excl. failed tests†  | Margins  | Full sample | Excl. failed tests†  | Margins  |
| Robot            | -0.710     | -0.874        | -0.177   | -0.713     | -0.931        | -0.177   |
|                  | (0.378)*   | (0.408)**     | (0.094)* | (0.387)*   | (0.418)**     | (0.0963)*|
| Cooperate (Stage 1) | 1.131     | 1.012        | 0.281    | 0.429***   | 0.453**       | 0.106***  |
|                  | (0.12)     | 0.13         |          | (0.457)**   | (0.524)**     | (0.0907)**|
| Inst. order      | Yes       | Yes          | Yes      | Yes       | Yes          | Yes      |
| Wave             | Yes       | Yes          | Yes      | Yes       | Yes          | Yes      |
| Exp. day         | Yes       | Yes          | Yes      | Yes       | Yes          | Yes      |
| Pseudo R-sq.     | 0.08      | 0.10         |          | 0.12      | 0.13         |          |
| Obs              | 148       | 127          | 148      | 148       | 127          | 148      |
| LL               | -94       | -79          | -90      | -90       | -76          | -76      |
| AIC              | 216       | 185          | 210      | 182       |
| BIC              | 257       | 225          | 255      | 225       |

| DVR condition | Benchmark | Including Choice at Stage 1 | Benchmark | Including Choice at Stage 1 |
|----------------|-----------|----------------------------|-----------|----------------------------|
|                  | Full sample | Excl. failed tests†  | Margins  | Full sample | Excl. failed tests†  | Margins  |
| Robot            | -0.504     | -0.482        | -0.106   | -0.842     | -0.892        | -0.170   |
|                  | (0.380)    | (0.415)       | (0.080)  | (0.424)**  | (0.471)*      | (0.085)**|
| Cooperate (Stage 1) | 1.885     | 2.253        | 0.381    | 0.457***   | 0.524***      | 0.0907***|
|                  | (0.17)     | 0.22         |          | (0.097)**  |             |          |
| Inst. order      | Yes       | Yes          | Yes      | Yes       | Yes          | Yes      |
| Wave             | Yes       | Yes          | Yes      | Yes       | Yes          | Yes      |
| Exp. day         | Yes       | Yes          | Yes      | Yes       | Yes          | Yes      |
| Pseudo R-sq.     | 0.07      | 0.09         |          | 0.17      | 0.22         |          |
| Obs              | 157       | 141          | 157      | 157       | 141          | 157      |
| LL               | -91       | -81          | -82      | -82       | -69          | -69      |
| AIC              | 210       | 189          | 193      | 193       | 169          | 169      |
| BIC              | 253       | 230          | 239      | 239       | 213          | 213      |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses; bootstrap standard errors between brackets.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11 provides a test of H2, directly addressing the effect of verbal reactions performed by both human and robot partners. Also in this case, our test is performed through a sample split, in this case by partner types: by investigating the effect of the DVR within the same experimental condition (i.e. the partner type) we are able to investigate the pure effect of verbal interactions irrespective of partner type.

As the Table shows, the coefficient of DVR is strongly positive and significant in both the Human and Robot experimental conditions. The upper panel in Table 11 shows that when the robot partner performs a DVR, subjects are on average 27% more likely to
Table 11: Effect of Treatment, by partner type, experimental sample

| Partner=Robot | Benchmark | Including Choice at Stage 1 |
|---------------|-----------|----------------------------|
|               | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
| DVR           | 1.073      | 1.149               | 0.268   | 1.089       | 1.160               | 0.272   |
|               | (0.374)*** | (0.391)***          | (0.0933)*** | (0.394)*** | (0.411)***          | (0.0983)*** |
| Cooperate (Stage 1) | 1.570      | 1.512               | 0.392   | 1.556       | 1.525               | 0.392   |
|               | (0.456)*** | (0.481)***          | (0.114)*** | (0.465)*** | (0.493)***          | (0.114)*** |
| Inst. order   | Yes        | Yes                 | Yes     | Yes         | Yes                 | Yes     |
| Wave          | Yes        | Yes                 | Yes     | Yes         | Yes                 | Yes     |
| Exp. day      | Yes        | Yes                 | Yes     | Yes         | Yes                 | Yes     |
| Pseudo R-sq.  | 0.14       | 0.15                |         | 0.20        | 0.21                |         |
| Obs           | 152        | 140                 | 152     | 152         | 140                 | 152     |
| LL            | -91        | -82                 |         | -84         | -77                 |         |
| AIC           | 210        | 192                 |         | 198         | 183                 |         |
| BIC           | 252        | 234                 |         | 244         | 227                 |         |

| Partner=Human | Benchmark | Including Choice at Stage 1 |
|---------------|-----------|----------------------------|
|               | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
| DVR           | 0.937      | 0.977               | 0.209   | 1.131       | 1.130               | 0.246   |
|               | (0.379)**  | (0.425)**           | (0.0840)** | (0.411)*** | (0.465)***          | (0.0881)*** |
| Cooperate (Stage 1) | 1.598      | 1.798               | 0.348   | 1.556       | 1.798               | 0.348   |
|               | (0.433)*** | (0.478)***          | (0.0930)*** | (0.465)*** | (0.493)***          | (0.114)*** |
| Inst. order   | Yes        | Yes                 | Yes     | Yes         | Yes                 | Yes     |
| Wave          | Yes        | Yes                 | Yes     | Yes         | Yes                 | Yes     |
| Exp. day      | Yes        | Yes                 | Yes     | Yes         | Yes                 | Yes     |
| Pseudo R-sq.  | 0.08       | 0.07                |         | 0.15        | 0.17                |         |
| Obs           | 153        | 128                 | 153     | 153         | 128                 | 153     |
| LL            | -92        | -77                 |         | -84         | -69                 |         |
| AIC           | 212        | 182                 |         | 199         | 168                 |         |
| BIC           | 254        | 222                 |         | 244         | 211                 |         |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses. † This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
choose Cooperate in the PD than in the control group. A similar result can be found in the Human subsample, in which the probability of cooperation is between 21% and 25% higher when the human partner performs a DVR than in the case he/she is not.\textsuperscript{29} It’s worth noting that the effect is particularly strong when controlling for the outcome of Stage 1, i.e. the choice of respondents in the previous “blind” stage of the experiment. In other words, the result holds when taking account of the subject’s cooperative attitude, which, as expected, is the most important predictor of her choices.

\textsuperscript{29}\textsuperscript{We also checked that this result is not affected by some sort of gender bias and found that no differential effect can be found depending on both partner’s gender, subject’s gender and the interaction between the two. A summary of this robustness check is reported in the Appendix, in Table 9, showing substantially no significant coefficients in the interaction terms.}
7 Conclusions

In a seminal paper, Farrel and Rabin state: “People don’t usually take the destructively agnostic attitude that ‘I won’t presume that the words mean what they have always meant’. Rather, people take the usual or literal meaning seriously” (Farrell and Rabin, 1996, p. 108). What is true for Human-Human interactions, does it hold for Human-Robot interactions?

In this paper, we devised a randomized experiment in which human subjects are randomly matched to either a human or an anthropomorphic robot partner and asked to perform a repeated Prisoner’s Dilemma to investigate (i) whether subjects behave differently depending on the nature of their partner (human or robot); (ii) whether a Dialogic Verbal Reaction (DVR), implicitly referring to cooperation as a socially desirable strategy, influenced the following subject’s choice; (iii) whether the effect played by DVR depended on the (human or robotic) nature of the partner.

Our results suggest that subjects tend to act more cooperatively with fellow humans, rather than with robots; are influenced by a DVR towards a more cooperative strategy; finally, that the effect of DVR is strong enough to make the difference in behavior, based on the partner’s type, insignificant. These results may suggest an interesting implication, i.e. the possibility to extend to humanoid robots the definition of “people” used by Farrel and Rabin in the quote above.

This last result may depend on the belief that a robot who is able to implement a dialogical interaction with the subject appears as “more human”, thus deserving a behavior similar to that reserved to fellow human partners. Further research is still needed in order to disentangle the effect of verbal interactions in HRI versus the social desirability of the message.

References

Ai, C. and E. C. Norton (2003). Interaction terms in logit and probit models. *Economics letters* 80(1), 123–129.

Balliet, D. (2010). Communication and cooperation in social dilemmas: A meta-analytic review. *Journal of Conflict Resolution* 54(1), 39–57.
Bicchieri, C. (2002). Covenants without swords: Group identity, norms, and communication in social dilemmas. *Rationality and Society* 14(2), 192–228.

Bicchieri, C. and A. Lev-On (2007). Computer-mediated communication and cooperation in social dilemmas: an experimental analysis. *Politics, Philosophy & Economics* 6(2), 139–168.

Biocca, F., C. Harms, and J. K. Burgoon (2003). Toward a more robust theory and measure of social presence: Review and suggested criteria. *Presence: Teleoperators & Virtual Environments* 12(5), 456–480.

Braver, S. L. and L. Wilson (1986). Choices in social dilemmas: Effects of communication within subgroups. *Journal of Conflict Resolution* 30(1), 51–62.

Buber, M. (1958). *I and Thou*. Scribners.

Burks, S. V., J. P. Carpenter, L. Goette, and A. Rustichini (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences* 106(19), 7745–7750.

Carter, J. R. and M. D. Irons (1991). Are economists different, and if so, why? *Journal of Economic Perspectives* 5(2), 171–177.

Compagni, A., V. Mele, and D. Ravasi (2015). How early implementations influence later adoptions of innovation: Social positioning and skill reproduction in the diffusion of robotic surgery. *Academy of Management Journal* 58(1), 242–278.

Conde-Valverde, M., I. Martínez, R. M. Quam, M. Rosa, A. D. Velez, C. Lorenzo, P. Jarabo, J. M. B. de Castro, E. Carbonell, and J. L. Arsuaga (2021). Neanderthals and homo sapiens had similar auditory and speech capacities. *Nature Ecology & Evolution*, 1–7.

Cook, K. S. and R. M. Cooper (2003). Experimental studies of cooperation, trust, and social exchange.

Crandall, J. W., M. Oudah, F. Ishowo-Oloko, S. Abdallah, J.-F. Bonnefon, M. Cebrian, A. Shariff, M. A. Goodrich, I. Rahwan, et al. (2018). Cooperating with machines. *Nature Communications* 9(1), 1–12.
Damiano, L. and P. Dumouchel (2018). Anthropomorphism in human–robot co-evolution. *Frontiers in psychology* 9, 468.

Dasgupta, P. (1988). Trust as a commodity, 49-72.

Dawes, R. M., J. McTavish, and H. Shaklee (1977). Behavior, communication, and assumptions about other people’s behavior in a commons dilemma situation. *Journal of personality and social psychology* 35(1), 1.

de Melo, C. M., P. Carnevale, and J. Gratch (2011). The effect of expression of anger and happiness in computer agents on negotiations with humans. In *The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pp. 937–944.

de Melo, C. M. and K. Terada (2020). The interplay of emotion expressions and strategy in promoting cooperation in the iterated prisoner’s dilemma. *Scientific reports* 10(1), 1–8.

DeAngelo, G. and B. C. McCannon (2020). Psychological game theory in public choice. *Public Choice* 182(1-2), 159–180.

DeSteno, D., C. Breazeal, R. H. Frank, D. Pizarro, J. Baumann, L. Dickens, and J. J. Lee (2012). Detecting the trustworthiness of novel partners in economic exchange. *Psychological science* 23(12), 1549–1556.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011). Individual risk attitudes: Measurement, determinants, and behavioral consequences. *Journal of the european economic association* 9(3), 522–550.

Dolbear Jr, F. T. and L. B. Lave (1966). Risk orientation as a predictor in the prisoner’s dilemma. *Journal of Conflict Resolution* 10(4), 506–515.

Dufwenberg, M. and G. Kirchsteiger (2004). A theory of sequential reciprocity. *Games and economic behavior* 47(2), 268–298.

Dumouchel, P. and L. Damiano (2017). *Living with robots*. Cambridge, USA: Harvard University Press.

Dunbar, R. and D. Bickerton (1996). Grooming, gossip and the evolution of language. *Nature* 380(6572), 303–303.
Farrell, J. and M. Rabin (1996). Cheap talk. *The Journal of Economic Perspectives* 10(3), 103–118.

Fong, T., I. Nourbakhsh, and K. Dautenhahn (2003). A survey of socially interactive robots. *Robotics and autonomous systems* 42(3-4), 143–166.

Frank, R. H., T. Gilovich, and D. T. Regan (1993). Does studying economics inhibit cooperation? *Journal of economic perspectives* 7(2), 159–171.

Frey, B. S. and S. Meier (2005). Selfish and indoctrinated economists? *European Journal of Law and Economics* 19(2), 165–171.

Gaggioli, A., A. Chirico, D. Di Lernia, M. A. Maggioni, F. Manzi, A. Marchetti, D. Massaro, D. Rossignoli, G. Sandini, D. Villani, G. Riva, and A. Sciutti (2021). Relationship development with humanoid social robots: Applying interpersonal theories to human/robot interaction. *Cyberpsychology, Behavior, and Social Networking.*

Geanakoplos, J., D. Pearce, and E. Stacchetti (1989). Psychological games and sequential rationality. *Games and economic Behavior* 1(1), 60–79.

Gelin, R. (2018). *NAO*, pp. 1–22. Dordrecht: Springer Netherlands.

Habermas, J. (1979). *Communication and the Evolution of Society*, Volume 572. Boston: Beacon Press.

Haucap, J. and A. Müller (2014). Why are economists so different? nature, nurture and gender effects in a simple trust game. *Nature, Nurture and Gender Effects in a Simple Trust Game* (March 1, 2014).

Heerink, M., B. Kröse, V. Evers, and B. Wielinga (2008). The influence of social presence on acceptance of a companion robot by older people.

Horan, R. D., E. Bulte, and J. F. Shogren (2005). How trade saved humanity from biological exclusion: an economic theory of neanderthal extinction. *Journal of Economic Behavior & Organization* 58(1), 1 – 29.

Horan, R. D., E. H. Bulte, and J. F. Shogren (2008). Coevolution of human speech and trade. *Journal of Economic Growth* 13(4), 293–313.
Horstmann, A. C., N. Bock, E. Linhuber, J. M. Szczuka, C. Straßmann, and N. C. Krämer (2018). Do a robot’s social skills and its objection discourage interactants from switching the robot off? PloS one 13(7), e0201581.

Hu, Y.-A. and D.-Y. Liu (2003). Altruism versus egoism in human behavior of mixed motives: An experimental study. American Journal of Economics and Sociology 62(4), 677–705.

Inter-University Consortium for Political Research (1964). Election Study. Ann Arbor, Michigan: University of Michigan.

Ishowo-Oloko, F., J.-F. Bonnefon, Z. Soroye, J. Crandall, I. Rahwan, and T. Rahwan (2019). Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. Nature Machine Intelligence 1(11), 517–521.

Johnson, J. (1993). Is talk really cheap? prompting conversation between critical theory and rational choice. American Political Science Review, 74–86.

Jones, B. A. and H. Rachlin (2009). Delay, probability, and social discounting in a public goods game. Journal of the experimental analysis of behavior 91(1), 61–73.

Jung, E.-S., B.-K. Kim, and S.-Y. Lee (2012). Trust and equity theory in prisoner’s dilemma. In International Conference on Neural Information Processing, pp. 560–567. Springer.

Kiesler, S., L. Sproull, and K. Waters (1996). A prisoner’s dilemma experiment on cooperation with people and human-like computers. Journal of personality and social psychology 70(1), 47.

Körner, U., K. Müller-Thur, T. Lunau, N. Dragano, P. Angerer, and A. Buchner (2019). Perceived stress in human–machine interaction in modern manufacturing environments—results of a qualitative interview study. Stress and Health 35(2), 187–199.

Krach, S., F. Hegel, B. Wrede, G. Sagerer, F. Binkofski, and T. Kircher (2008). Can machines think? interaction and perspective taking with robots investigated via fmri. PloS one 3(7), e2597.
Kreps, D. M. (1990). *A Course on Microeconomic Theory*. Princeton University Press.

Kreps, D. M., P. Milgrom, J. Roberts, and R. Wilson (1982). Rational cooperation in the finitely repeated prisoners’ dilemma. *Journal of Economic Theory 27*(2), 245–252.

Laban, G., J.-N. George, V. Morrison, and E. S. Cross (2021). Tell me more! Assessing interactions with social robots from speech. *Paladyn, Journal of Behavioral Robotics 12*(1), 136–159.

Lieberman, P. (2006). *Toward an evolutionary biology of language*. Harvard University Press.

Locey, M. L., V. Safin, and H. Rachlin (2013). Social discounting and the prisoner’s dilemma game. *Journal of the experimental analysis of behavior 99*(1), 85–97.

Malesza, M. (2020). The effects of the dark triad traits in prisoner’s dilemma game. *Current Psychology 39*(3), 1055–1062.

Manzi, F., C. Di Dio, D. Di Lernia, D. Rossignoli, M. A. Maggioni, D. Massaro, A. Marchetti, and G. Riva (2021). Can you activate me? from robots to human brain. *Frontiers in Robotics and AI 8*.

March, C. (2019). *The behavioral economics of artificial intelligence: Lessons from experiments with computer players*. Number 154. BERG Working Paper Series.

Marchetti, A., F. Manzi, S. Itakura, and D. Massaro (2018). Theory of mind and humanoid robots from a lifespan perspective. *Zeitschrift für Psychologie 226*(2), 98.

Mazar, N., O. Amir, and D. Ariely (2008). The dishonesty of honest people: A theory of self-concept maintenance. *Journal of Marketing Research 45*(6), 633–644.

McCabe, D. L. and L. K. Trevino (1993). Academic dishonesty: Honor codes and other contextual influences. *The Journal of Higher Education 64*(5), 522–538.

McCabe, K., D. Houser, L. Ryan, V. Smith, and T. Trouard (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences 98*(20), 11832–11835.
Miwa, K. and H. Terai (2012). Impact of two types of partner, perceived or actual, in human–human and human–agent interaction. *Computers in Human Behavior* 28(4), 1286–1297.

Nouri, E. and D. Traum (2013). A cross-cultural study of playing simple economic games online with humans and virtual humans. In *International Conference on Human-Computer Interaction*, pp. 266–275. Springer.

Ostrom, E. (2000). Collective action and the evolution of social norms. *Journal of economic perspectives* 14(3), 137–158.

Ostrom, E. and J. Walker (1991). Communication in a commons: cooperation without external enforcement. *Laboratory research in political economy*, 287–322.

Ostrom, E. and J. Walker (1997). Neither markets nor states: Linking transformation process in collective action arenas. In D. C. Mueller (Ed.), *Perspectives on public choice: A handbook*, pp. 35–72. Cambridge: Cambridge University Press.

Paeng, E., J. Wu, and J. Boerkoel (2016). Human-robot trust and cooperation through a game theoretic framework. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Volume 30.

Proto, E., A. Rustichini, and A. Sofianos (2019). Intelligence, personality, and gains from cooperation in repeated interactions. *Journal of Political Economy* 127(3), 1351–1390.

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American economic review*, 1281–1302.

Rilling, J. K., D. A. Gutman, T. R. Zeh, G. Pagnoni, G. S. Berns, and C. D. Kilts (2002). A neural basis for social cooperation. *Neuron* 35(2), 395–405.

Rilling, J. K., A. G. Sanfey, J. A. Aronson, L. E. Nystrom, and J. D. Cohen (2004). The neural correlates of theory of mind within interpersonal interactions. *Neuroimage* 22(4), 1694–1703.

Robaczewski, A., J. Bouchard, K. Bouchard, and S. Gaboury (2020). Socially assistive robots: The specific case of the nao. *International Journal of Social Robotics*, 1–37.
Robelski, S. and S. Wischniewski (2018). Human-machine interaction and health at work: a scoping review. *International journal of human factors and ergonomics* 5(2), 93–110.

Sabater-Grande, G. and N. Georgantzis (2002). Accounting for risk aversion in repeated prisoners’ dilemma games: An experimental test. *Journal of economic behavior & organization* 48(1), 37–50.

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and society* 7(1), 58–92.

Skeath, A. D.-S. (1999). *Games of Strategy.* ed. WW Norton & Company.

Streich, P. and J. S. Levy (2007). Time horizons, discounting, and intertemporal choice. *Journal of Conflict Resolution* 51(2), 199–226.

Sung, J., R. E. Grinter, and H. I. Christensen (2010). Domestic robot ecology. *International Journal of Social Robotics* 2(4), 417–429.

Tattersall, I. (2008). An evolutionary framework for the acquisition of symbolic cognition by homo sapiens. *Comparative cognition & behavior reviews* 3.

Tedeschi, J. T., D. S. Hiester, and J. P. Gahagan (1969). Trust and the prisoner’s dilemma game. *The Journal of Social Psychology* 79(1), 43–50.

Terada, K. and C. Takeuchi (2017). Emotional expression in simple line drawings of a robot’s face leads to higher offers in the ultimatum game. *Frontiers in psychology* 8, 724.

Thaler, R. (1981). Some empirical evidence on dynamic inconsistency. *Economics letters* 8(3), 201–207.

Turing, A. M. (1950). Computing machinery and intelligence. *Mind* 59(236), 433.

Wu, J., E. Paeng, K. Linder, P. Valdesolo, and J. C. Boerkoel (2016). Trust and cooperation in human-robot decision making. In *2016 aaai fall symposium series*.

Yezer, A. M., R. S. Goldfarb, and P. J. Poppen (1996). Does studying economics discourage cooperation? watch what we do, not what we say or how we play. *Journal of Economic Perspectives* 10(1), 177–186.
Yi, R., M. W. Johnson, and W. K. Bickel (2005). Relationship between cooperation in an iterated prisoner’s dilemma game and the discounting of hypothetical outcomes. *Learning & behavior* 33(3), 324–336.
Appendix A  Linear Probability Models (OLS)

This section presents a replication of the main results shown in Section 5 by adopting a Linear Probability Model, estimates through Ordinary Least Squares (OLS), in place of Logit models adopted in the main texts.
Table A1: LPM: Effect of Partner type, by treatment condition, experimental sample

| No DVR condition | Benchmark | Including Choice at Stage 1 |
|------------------|-----------|----------------------------|
|                  | Full sample | Excl. failed tests\[\dagger\] | Full sample | Excl. failed tests\[\dagger\] |
| Robot            | -0.159     | (0.091)*                   | -0.151     | (0.091)*                   |
|                  | -0.196     | (0.097)**                  | -0.200     | (0.095)**                  |
| Cooperate (Stage 1) |            |                            | 0.233     | (0.088)***                 |
|                  |            |                            | 0.209     | (0.094)**                  |
| Inst. order      | Yes        | Yes                        | Yes       | Yes                        |
| Wave             | Yes        | Yes                        | Yes       | Yes                        |
| Exp. day         | Yes        | Yes                        | Yes       | Yes                        |
| Adj. R-sq.       | 0.02       | 0.03                       | 0.06      | 0.06                       |
| Obs              | 148        | 127                        | 148       | 127                        |
| LL               | -99        | -83                        | -95       | -80                        |
| AIC              | 225        | 194                        | 220       | 190                        |
| BIC              | 267        | 233                        | 265       | 233                        |

| DVR condition | Benchmark | Including Choice at Stage 1 |
|---------------|-----------|----------------------------|
|                | Full sample | Excl. failed tests\[\dagger\] | Full sample | Excl. failed tests\[\dagger\] |
| Robot          | -0.100     | (0.080)                     | -0.143     | (0.078)*                   |
|                | -0.098     | (0.085)                     | -0.151     | (0.081)*                   |
| Cooperate (Stage 1) |            |                            | 0.359     | (0.086)***                 |
|                |            |                            | 0.407     | (0.089)***                 |
| Inst. order    | Yes        | Yes                        | Yes       | Yes                        |
| Wave           | Yes        | Yes                        | Yes       | Yes                        |
| Exp. day       | Yes        | Yes                        | Yes       | Yes                        |
| Adj. R-sq.     | 0.01       | 0.03                       | 0.12      | 0.17                       |
| Obs            | 157        | 141                        | 157       | 141                        |
| LL             | -96        | -85                        | -85       | -73                        |
| AIC            | 219        | 197                        | 201       | 175                        |
| BIC            | 262        | 238                        | 247       | 219                        |

Notes. LPM model (OLS), dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Robust standard errors between parentheses.  
\[\dagger\] This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.  
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table A2: LPM: Effect of Treatment, by partner type, experimental sample

|                  | Benchmark | Including Choice at Stage 1 |
|------------------|-----------|----------------------------|
|                  | Full sample | Excl. failed tests$^\dagger$ | Full sample | Excl. failed tests$^\dagger$ |
| **Partner=Robot**|            |                            |            |                            |
| Choice: “Cooperate” |           |                             |            |                            |
| DVR              | 0.228      | 0.243                       | 0.211      | 0.224                       |
|                  | (0.082)$^{***}$ | (0.085)$^{***}$              | (0.078)$^{***}$ | (0.081)$^{***}$              |
| Cooperate (Stage 1) |            |                             | 0.308      | 0.286                       |
|                  |            |                             |            | (0.087)$^{***}$              | (0.091)$^{***}$              |
| Inst. order      | Yes        | Yes                         | Yes        | Yes                         |
| Wave             | Yes        | Yes                         | Yes        | Yes                         |
| Exp. day         | Yes        | Yes                         | Yes        | Yes                         |
| Adj. R-sq.       | 0.10       | 0.11                        | 0.17       | 0.17                        |
| Obs              | 152        | 140                         | 152        | 140                         |
| LL               | -96        | -87                         | -89        | -81                         |
| AIC              | 219        | 201                         | 208        | 192                         |
| BIC              | 261        | 242                         | 253        | 236                         |
| **Partner=Human**|            |                             |            |                            |
| Choice: “Cooperate” |           |                             |            |                            |
| DVR              | 0.200      | 0.206                       | 0.207      | 0.205                       |
|                  | (0.080)$^{**}$ | (0.089)$^{**}$              | (0.079)$^{***}$ | (0.088)$^{**}$              |
| Cooperate (Stage 1) |            |                             | 0.313      | 0.352                       |
|                  |            |                             |            | (0.086)$^{***}$              | (0.091)$^{***}$              |
| Inst. order      | Yes        | Yes                         | Yes        | Yes                         |
| Wave             | Yes        | Yes                         | Yes        | Yes                         |
| Exp. day         | Yes        | Yes                         | Yes        | Yes                         |
| Adj. R-sq.       | 0.01       | -0.01                       | 0.10       | 0.10                        |
| Obs              | 153        | 128                         | 153        | 128                         |
| LL               | -97        | -81                         | -89        | -73                         |
| AIC              | 221        | 190                         | 208        | 175                         |
| BIC              | 264        | 229                         | 254        | 218                         |

Notes. LPM model (OLS), dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Robust standard errors between parentheses.

$^\dagger$ This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|---------------------------|
|                     | Full sample | Excl. failed tests† | Full sample | Excl. failed tests† |
| DVR × Robot         | 0.043      | 0.0843                   | 0.016      | 0.066              |
|                     | (0.114)    | (0.123)                  | (0.111)    | (0.118)             |
| DVR                 | 0.184      | 0.166                    | 0.184      | 0.156              |
|                     | (0.081)**  | (0.089)*                 | (0.079)**  | (0.088)*            |
| Robot               | -0.162     | -0.206                   | -0.167     | -0.224             |
|                     | (0.085)*   | (0.091)**                | (0.084)**  | (0.090)**           |
| Cooperate (Stage 1) |            |                          | 0.283      | 0.297              |
|                     |            |                          | (0.056)**  | (0.063)**           |
| Inst. order         | Yes        | Yes                      | Yes        | Yes                |
| Wave                | Yes        | Yes                      | Yes        | Yes                |
| Exp. day            | Yes        | Yes                      | Yes        | Yes                |
| Adj. R-sq.          | 0.05       | 0.06                     | 0.12       | 0.13               |
| Obs                 | 305        | 268                      | 305        | 268                |
| LL                  | -202       | -176                     | -190       | -164               |
| AIC                 | 435        | 383                      | 414        | 362                |
| BIC                 | 495        | 441                      | 477        | 423                |

Notes. LPM model (OLS), dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Robust standard errors between parentheses.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$
Table A4: LPM, Robustness check: excluding Econ students

| Choice: “Cooperate” | Benchmark Including Choice at Stage 1 | Excl. failed tests | Excl. failed tests† |
|---------------------|-------------------------------------|--------------------|--------------------|
|                     | Full sample                         | Excl. failed tests | Full sample        | Excl. failed tests† |
| DVR × Robot         | 0.034                               | 0.124              | 0.028              | 0.138              |
|                     | (0.145)                             | (0.156)            | (0.140)            | (0.152)            |
| DVR                 | 0.217                               | 0.179              | 0.199              | 0.143              |
|                     | (0.094)**                           | (0.108)*           | (0.095)**          | (0.109)            |
| Robot               | -0.147                              | -0.210             | -0.173             | -0.262             |
|                     | (0.114)                             | (0.123)*           | (0.114)            | (0.123)**          |
| Cooperate (Stage 1) |                                     | 0.264              | 0.287              |
|                     |                                     | (0.080)**          | (0.088)**          |

| Inst. order         | Yes                                 | Yes                | Yes                | Yes                |
| Wave                | Yes                                 | Yes                | Yes                | Yes                |
| Exp. day            | Yes                                 | Yes                | Yes                | Yes                |
| Adj. R-sq.          | 0.08                                | 0.09               | 0.13               | 0.15               |
| Obs                 | 192                                 | 166                | 192                | 166                |
| LL                  | -118                                | -100               | -112               | -94                |
| AIC                 | 269                                 | 233                | 259                | 223                |
| BIC                 | 321                                 | 283                | 314                | 275                |

Notes. LPM model (OLS), dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Students attending BAs in Economics, Management, Finance and Banking are excluded from the sample. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

\* p < 0.10, ** p < 0.05, *** p < 0.01
Table A5: LPM, Robustness check: controlling for psychological and behavioral factors

| Choice: “Cooperate” | Benchmark Including Choice at Stage 1 | Full sample | Excl. failed tests† | Full sample | Excl. failed tests† |
|---------------------|----------------------------------------|--------------|----------------------|--------------|----------------------|
| DVR×Robot           |                                        | 0.053        | 0.095                | 0.023        | 0.072                |
|                     |                                        | (0.115)      | (0.123)              | (0.112)      | (0.119)              |
| DVR                 |                                        | 0.176        | 0.157                | 0.179        | 0.152                |
|                     |                                        | (0.081)**     | (0.089)*             | (0.079)**     | (0.088)*             |
| Robot               |                                        | -0.170       | -0.212               | -0.172       | -0.227               |
|                     |                                        | (0.086)**     | (0.092)**            | (0.085)**     | (0.091)**            |
| Psychological and behavioral controls: | | | | | |
| Risk scale          |                                        | 0.001        | 0.002                | -0.003       | -0.002               |
|                     |                                        | (0.013)      | (0.014)              | (0.012)      | (0.013)              |
| Trust scale         |                                        | 0.048        | 0.041                | 0.041        | 0.036                |
|                     |                                        | (0.029)      | (0.032)              | (0.028)      | (0.031)              |
| IDR                 |                                        | -0.001       | -0.001               | -0.001       | -0.000               |
|                     |                                        | (0.001)      | (0.001)              | (0.001)      | (0.001)              |
| Donation            |                                        | 0.037        | 0.015                | 0.027        | 0.008                |
|                     |                                        | (0.058)      | (0.063)              | (0.056)      | (0.060)              |
| Cooperate (Stage 1) |                                        |              | 0.276                | 0.293        |
|                     |                                        |              | (0.061)**            | (0.065)***** |
| Inst. order         | Yes                                    | Yes          | Yes                  | Yes          | Yes                  |
| Wave                | Yes                                    | Yes          | Yes                  | Yes          | Yes                  |
| Exp. day            | Yes                                    | Yes          | Yes                  | Yes          | Yes                  |
| Adj. R-sq.          | 0.05                                   | 0.05         | 0.11                 | 0.13         |
| Obs                 | 305                                    | 268          | 305                  | 268          |
| LL                  | -200                                   | -175         | -189                 | -163         |
| AIC                 | 440                                    | 389          | 420                  | 369          |
| BIC                 | 514                                    | 461          | 498                  | 444          |

Notes. LPM model (OLS), dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses.
† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.
* p < 0.10, ** p < 0.05, *** p < 0.01
Table A6: LPM, Robustness check: controlling for heterogeneity driven by perception about robot’s behavior

| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|----------------------------|
|                     | Full sample | Excl. failed tests† | Full sample | Excl. failed tests† |
| DVR×Adaptive        | -0.165     | -0.189                   | -0.185     | -0.197              |
|                     | (0.167)    | (0.176)                  | (0.159)    | (0.168)             |
| DVR                 | 0.298      | 0.325                    | 0.290      | 0.309               |
|                     | (0.105)*** | (0.109)***               | (0.101)*** | (0.105)***          |
| Adaptive            | 0.142      | 0.139                    | 0.134      | 0.129               |
|                     | (0.127)    | (0.131)                  | (0.124)    | (0.130)             |
| Cooperate (Stage 1) |            |                          | 0.309      | 0.286               |
|                     |            |                          | (0.087)*** | (0.090)***          |
| Inst. order         | Yes        | Yes                      | Yes        | Yes                 |
| Wave                | Yes        | Yes                      | Yes        | Yes                 |
| Exp. day            | Yes        | Yes                      | Yes        | Yes                 |
| Adj. R-sq.          | 0.09       | 0.11                     | 0.17       | 0.17                |
| Obs                 | 152        | 140                      | 152        | 140                 |
| LL                  | -95        | -86                      | -88        | -80                 |
| AIC                 | 221        | 203                      | 210        | 194                 |
| BIC                 | 270        | 251                      | 261        | 244                 |

Notes. LPM model (OLS), dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses.
† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.
* p < 0.10, ** p < 0.05, *** p < 0.01
| Choice: “Cooperate”                        | Benchmark (Full sample) | Benchmark (Excl. failed tests) | Including Choice at Stage 1 (Full sample) | Including Choice at Stage 1 (Excl. failed tests) |
|------------------------------------------|-------------------------|--------------------------------|-------------------------------------------|------------------------------------------------|
| DVR × Female × Female partner            | -0.450 (0.364)          | -0.343 (0.419)                | -0.233 (0.360)                            | -0.062 (0.415)                               |
| DVR × Female                             | 0.066 (0.292)           | 0.025 (0.353)                 | -0.049 (0.284)                            | -0.152 (0.345)                               |
| DVR × Female partner                     | 0.324 (0.287)           | 0.226 (0.323)                 | 0.172 (0.286)                             | 0.027 (0.325)                                |
| Female × Female partner                  | 0.518 (0.282)*          | 0.437 (0.325)                 | 0.402 (0.283)                             | 0.280 (0.335)                                |
| DVR                                      | 0.132 (0.226)           | 0.178 (0.268)                 | 0.218 (0.222)                             | 0.301 (0.269)                                |
| Female                                   | -0.282 (0.230)          | -0.237 (0.270)                | -0.239 (0.227)                            | -0.152 (0.276)                               |
| Female partner                           | -0.414 (0.245)*         | -0.272 (0.289)                | -0.331 (0.244)                            | -0.178 (0.292)                               |
| Cooperate (Stage 1)                      |                         |                                | 0.309 (0.086)***                         | 0.353 (0.092)***                             |

**Inst. order** | Yes | Yes | Yes | Yes |
| Wave          | Yes | Yes | Yes | Yes |
| Exp. day      | Yes | Yes | Yes | Yes |

| Adj. R-sq. | Obs | LL | AIC | BIC |
|------------|-----|----|-----|-----|
| 0.01       | 153 | -93| 227 | 287 |
| -0.03      | 128 | -79| 197 | 254 |
| 0.09       | 153 | -86| 214 | 277 |
| 0.09       | 128 | -70| 183 | 242 |

**Notes.** Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses. Control variables include: Female, Freshman, Economics and Failed test.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* p < 0.10, ** p < 0.05, *** p < 0.01

Electronic copy available at: https://ssrn.com/abstract=3834619
Appendix B  Balance tests

Figure B1: Summary of balance tests for assignment to treatment (Dialog)
Figure B2: Summary of balance tests for assignment to Human/Robot condition
Balance tests for experimental wave, 1st year students
Test: 2021 vs 2020

Figure B3: Summary of balance tests for experimental wave
Appendix C Additional tables

This section presents additional tables that are referenced in the main text but moved here for reasons of space.

Table C1: Cooperation patterns at Stage 2, by Treatment group and Experimental condition

| Treatment Group    | Mean | St. Err. | Obs |
|--------------------|------|----------|-----|
| **No DVR group**   |      |          |     |
| Partner=Robot      | 0.380| 0.058    | 71  |
| Partner=Human      | 0.558| 0.057    | 77  |
| **DVR group**      |      |          |     |
| Partner=Robot      | 0.630| 0.054    | 81  |
| Partner=Human      | 0.737| 0.051    | 76  |

Electronic copy available at: https://ssrn.com/abstract=3834619
Table C2: Outline of *Dialogic Verbal Reactions* performed by Partner

| Outcome (Subject, Partner) | DVR type | Speech performed by partner |
|----------------------------|----------|-----------------------------|
| (C, N)                     | Apology  | “I realized I made a mistake on our online interaction. I wanted to press C to cooperate. However I pressed N by mistake. I am really sorry! I will be more careful next time” |
| (N, C)                     | Reprimand| “I am really upset. If you had chosen to cooperate we would have gained 6 euro each, a reasonable amount! On the contrary you exploited my goodwill and I got nothing” |
| (N, N)                     | Disappointment | “What a pity! If we had choose to cooperate we would have gained 6 euros each. Why not cooperating in the next round?” |
| (C, C)                     | None     |                             |
| **Outcome variables** | Description |
|----------------------|-------------|
| Choice (Stage 2)     | Main outcome, binary variable equal to 1 if subject chooses “Cooperate” in PD’s Stage 2, 0 otherwise |
| Choice (Stage 1)     | Binary variable equal to 1 if subject chooses “Cooperate” in PD’s Stage 1, 0 otherwise |

| **Treatment variables** | Description |
|-------------------------|-------------|
| DVR                     | Treatment variable, assuming value equal to 1 if partner performs a Dialogic Verbal Reaction after the outcome of Stage 1 |
| Robot                   | Experimental condition, assuming value equal to 1 if subject is assigned to a Robot partner, 0 if he/she is assigned to a Human partner |

| **Experiment-related controls** | Description |
|-------------------------------|-------------|
| Instruction order            | Categorical variable recording the PD’s outcome shown first to subject during the instruction session. Categories: “CC” (omitted reference category), “CN”, “NC”, “NN”, where first letter refers to subjects’ choice, second to partners’ |
| Wave                         | Binary variable, equal to 1 if experiment occurred in 2021, 0 in 2020 |
| Experimental day             | Categorical variable recording the sequential number of each day of the experiment. Values from 1 (omitted reference category) for first day to 9 for ninth day |

| **Individual characteristics** | Description |
|-------------------------------|-------------|
| Female                        | Binary variable, equal to 1 if subject is Female, 0 otherwise |
| Freshman                      | Binary variable, equal to 1 if subject is attending the first year of his/her BA, 0 if he/she is attending second year |
| Econ                          | Binary variable, equal to 1 if subject is enrolled in BAs in Economics, Management, Finance and Banking |
| Fail test                     | Binary variable, equal to 1 if subject failed the pre-game test after the instruction session |

| **Psychological scales and behaviors** | Description |
|--------------------------------------|-------------|
| Risk scale                           | General risk scale (Dohmen et al., 2011) |
| Trust scale                          | General trust scale (Inter-University Consortium for Political Research, 1964) |
| IDR                                  | Inter-temporal discount rate, proxied by a question comparing a hypothetical immediate reward of €10 to a large delayed amount (larger values in IDR implies higher discount rates) |
| Donation                             | Binary variable, equal to 1 if the subject choosed to donate 1 euro out of the final reward to a Charity of his/her choice, choosing between Medécins Sans Frontières or Greenpeace |

| **Other controls** | Description |
|--------------------|-------------|
| Adaptive           | Binary variable based on the answers given in the online questionnaire to the question: “When interacting with a human being how does a robot behave?” The variable is equal to 1 if the subject chose the following answer: “Robot adapts its behavior to the interaction with the human being”; the variable is equal to 0 if the subject chose one of the following answers: “Robot privileges its own interest to the interest of the human being”, “Robot privileges the interest of the human being to its own interest”, ”Robot follows a Random behavior”, Robots executes a fixed list of commands and operations” |
|                          | Mean  | St. Dev. | Obs  |
|--------------------------|-------|----------|------|
| **Experimental sample**  |       |          |      |
| **Outcome variables**    |       |          |      |
| Cooperate (Stage 1)      | 0.673 | 0.470    | 343  |
| Cooperate (Stage 2)      | 0.580 | 0.494    | 305  |
| **Control variables**    |       |          |      |
| Female                   | 0.697 | 0.460    | 343  |
| Freshman                 | 0.778 | 0.416    | 343  |
| Economics                | 0.397 | 0.490    | 343  |
| Test failed              | 0.128 | 0.335    | 343  |
| **Full sample**          |       |          |      |
| **Outcome variables**    |       |          |      |
| Cooperate (Stage 1)      | 0.793 | 0.406    | 541  |
| Cooperate (Stage 2)      | 0.659 | 0.474    | 490  |
| **Control variables**    |       |          |      |
| Female                   | 0.682 | 0.466    | 541  |
| Freshman                 | 0.786 | 0.411    | 541  |
| Economics                | 0.381 | 0.486    | 541  |
| Test failed              | 0.131 | 0.338    | 541  |

Notes: The Experimental sample includes only subjects whose outcomes in Stage 1 are relevant for Treatment/Control assignment, i.e. this sample excludes all subjects observing “CC” as Stage 1’s outcome; Full sample includes all subjects that participated to the research Phase 2.
Table C5: Effect of Partner type, by treatment condition, experimental sample, including control variables

| No DVR condition | Benchmark | Including Choice at Stage 1 |
|------------------|-----------|-----------------------------|
|                  | Full sample | Excl. failed tests\(^1\) | Margins | Full sample | Excl. failed tests\(^1\) | Margins |
| Robot            | -0.710     | -0.874                     | -0.177  | -0.713     | -0.931                     | -0.177  |
|                  | (0.378)*   | (0.408)**                  | (0.0941)* | (0.387)*   | (0.418)**                  | (0.0963)* |
| Cooperate (Stage 1) | 1.131     | 1.012                     | 0.281   | (0.429)**   | (0.453)**                  | (0.106)*** |
| Inst. order      | Yes        | Yes                       | Yes     | Yes        | Yes                       | Yes     |
| Wave             | Yes        | Yes                       | Yes     | Yes        | Yes                       | Yes     |
| Exp. day         | Yes        | Yes                       | Yes     | Yes        | Yes                       | Yes     |
| Pseudo R-sq.     | 0.08       | 0.10                      | 0.12    | 0.13       |                           |         |
| Obs              | 148        | 127                       | 148     | 148        | 127                       | 148     |
| LL               | -94        | -79                       | -90     | -76        |                           |         |
| AIC              | 216        | 185                       | 210     | 182        |                           |         |
| BIC              | 257        | 225                       | 255     | 225        |                           |         |

| DVR condition | Benchmark | Including Choice at Stage 1 |
|---------------|-----------|-----------------------------|
|                | Full sample | Excl. failed tests\(^1\) | Margins | Full sample | Excl. failed tests\(^1\) | Margins |
| Robot          | -0.504     | -0.482                     | -0.106  | -0.842     | -0.892                     | -0.170  |
|                | (0.380)    | (0.415)                    | (0.080) | (0.424)**  | (0.471)**                  | (0.085)** |
| Cooperate (Stage 1) | 1.885    | 2.253                     | 0.381   | (0.457)**   | (0.524)**                  | (0.0907)*** |
| Inst. order    | Yes        | Yes                       | Yes     | Yes        | Yes                       | Yes     |
| Wave           | Yes        | Yes                       | Yes     | Yes        | Yes                       | Yes     |
| Exp. day       | Yes        | Yes                       | Yes     | Yes        | Yes                       | Yes     |
| Pseudo R-sq.   | 0.07       | 0.09                      | 0.17    | 0.22       |                           |         |
| Obs            | 157        | 141                       | 157     | 157        | 141                       | 157     |
| LL             | -91        | -81                       | -82     | -60        |                           |         |
| AIC            | 210        | 189                       | 193     | 169        |                           |         |
| BIC            | 253        | 230                       | 239     | 213        |                           |         |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2. Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses. Control variables include: Female, Freshman, Economics and Failed test.

\(^1\) This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

\(* p < 0.10, \text{ } ** p < 0.05, \text{ } *** p < 0.01\)
Table C6: Heterogeneity of Treatment by partner type, experimental sample, including controls

| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|-----------------------------|
|                     | Full sample | Excl. failed tests† | Margins | Full sample | Excl. failed tests† | Margins |
| DVR×Robot           |            |                         |         |            |                         |         |
|                     | 0.134       | 0.357                    | 0.032   | 0.0170     | 0.300                    | 0.004   |
|                     | (0.520)     | (0.563)                  | (0.126) | (0.540)    | (0.586)                  | (0.130) |
| DVR                 | 0.832       | 0.737                    | 0.202   | 0.908      | 0.759                    | 0.219   |
|                     | (0.371)**   | (0.409)*                 | (0.087)**| (0.385)**  | (0.426)*                 | (0.093)**|
| Robot               | -0.643      | -0.859                   | -0.156  | -0.759     | -1.049                   | -0.183  |
|                     | (0.364)*    | (0.392)**                | (0.088)*| (0.377)**  | (0.412)**                | (0.091)**|
| Cooperate (Stage 1) |            |                         |         | 1.300      | 1.402                    | 0.314   |
|                     |             |                         |         | (0.292)**  | (0.316)**                | (0.071)***|
| Controls            | Yes         | Yes                      | Yes     | Yes        | Yes                      | Yes     |
| Inst. order         | Yes         | Yes                      | Yes     | Yes        | Yes                      | Yes     |
| Wave                | Yes         | Yes                      | Yes     | Yes        | Yes                      | Yes     |
| Exp. day            | Yes         | Yes                      | Yes     | Yes        | Yes                      | Yes     |
| Pseudo R-sq.        | 0.08        | 0.09                     | 0.13    | 0.15       |                         |         |
| Obs                 | 305         | 268                      | 305     | 305        | 268                      | 305     |
| LL                  | -191        | -166                     | -180    | -155       |                         |         |
| AIC                 | 422         | 370                      | 403     | 351        |                         |         |
| BIC                 | 496         | 438                      | 481     | 423        |                         |         |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses. Control variables include: Female, Freshman, Economics and Failed test.

† This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* p < 0.10, ** p < 0.05, *** p < 0.01
Table C7: Heterogeneity of Treatment by partner type, excluding Econ students, including control variables

| Choice: “Cooperate” | Benchmark | Including Choice at Stage 1 |
|---------------------|-----------|-----------------------------|
|                     | Full sample | Excl. failed tests\(^{†}\) | Margins | Full sample | Excl. failed tests\(^{†}\) | Margins |
| DVR × Robot         |            |                             |         |            |                             |         |
|                     | -0.004     | 0.453                       | -0.001  | -0.113     | 0.501                       | -0.026  |
|                     | (0.691)    | (0.756)                     | (0.159) | (0.717)    | (0.792)                     | (0.163) |
| DVR                 | 1.162      | 0.982                       | 0.268   | 1.193      | 0.885                       | 0.271   |
|                     | (0.484)**  | (0.540)*                    | (0.111)**| (0.507)**  | (0.569)                     | (0.113)**|
| Robot               | -0.650     | -0.938                      | -0.150  | -0.862     | -1.302                      | -0.196  |
|                     | (0.484)    | (0.529)*                    | (0.112) | (0.497)*   | (0.562)**                   | (0.113)*|
| Cooperate (Stage 1) |            |                             | 1.388   | 1.567      | 0.315                       |
|                     |            |                             |         | (0.416)*** | (0.468)***                  | (0.094)***|
| Controls            | Yes        | Yes                         | Yes     | Yes        | Yes                         | Yes     |
| Inst. order         | Yes        | Yes                         | Yes     | Yes        | Yes                         | Yes     |
| Wave                | Yes        | Yes                         | Yes     | Yes        | Yes                         | Yes     |
| Exp. day            | Yes        | Yes                         | Yes     | Yes        | Yes                         | Yes     |
| Pseudo R-sq.        | 0.13       | 0.14                        | 0.18    | 0.20       |
| Obs                 | 192        | 166                         | 192     | 192        |
| LL                  | -111       | -95                         | -105    | -89        |
| AIC                 | 261        | 226                         | 251     | 216        |
| BIC                 | 323        | 282                         | 316     | 275        |

Notes. Logit model, dependent variable: choice of Cooperation in stage 2, Prisoner’s Dilemma. Students attending BAs in Economics, Management, Finance and Banking are excluded from the sample. Subjects observing the outcome “CC” in Stage 1 are excluded from this sample. Standard errors between parentheses. \(^{†}\) This sub-sample excludes all subjects failing the test to assess their full comprehension of the instruction.

* \( p < 0.10 \), ** \( p < 0.05 \), *** \( p < 0.01 \)
Appendix D  Additional figures

Figure D1: Percentage of respondents being proposed to move to Stage 2 by observed outcome in Stage 1.
Figure D2: Percentage of respondents accepting to proceed to Stage 2 by observed outcome in Stage 1.
Figure D3: Choice of Cooperation in Stage 2 by experimental condition and type of DVR
Figure D4: NAO robot, produced by Softbank Robotics

Figure D5: Experiment room