International Journal of Social Robotics (2023) 15:791–805
https://doi.org/10.1007/s12369-023-00981-7

Human–Robot Cooperation in Economic Games: People Show Strong Reciprocity but Conditional Prosociality Toward Robots

Te-Yi Hsieh[1] · Bishakha Chaudhury[1] · Emily S. Cross[1,2,3]

Accepted: 15 February 2023 / Published online: 7 April 2023

Abstract

Understanding how people socially engage with robots is becoming increasingly important as these machines are deployed in social settings. We investigated 70 participants' situational cooperation tendencies towards a robot using prisoner's dilemma games, manipulating the incentives for cooperative decisions to be high or low. We predicted that people would cooperate more often with the robot in high-incentive conditions. We also administered subjective measures to explore the relationships between people's cooperative decisions and their social value orientation, attitudes towards robots, and anthropomorphism tendencies. Our results showed incentive structure did not predict human cooperation overall, but did influence cooperation in early rounds, where participants cooperated significantly more in high-incentive conditions. Exploratory analyses further revealed that participants played a tit-for-tat strategy against the robot (whose decisions were random), and only behaved prosocially toward the robot when they had achieved high scores themselves. These findings highlight how people make social decisions when their individual profit is at odds with collective profit with a robot, and advance understanding on human–robot interactions in collaborative contexts.

Keywords Human–robot interaction · Human–robot cooperation · Prisoner's dilemma games · Rapoport's K-index · Reciprocity

1 Introduction

Social robots are becoming valuable tools for assisting people with daily life, as they take on new roles in healthcare, education, and therapy [8]. However, many commercially available social robots suffer from the criticism of not fitting users' expectations, especially in terms of the richness or appropriateness of their social responses, which in turn diminishes people's ability to collaborate with these machines, let alone build long-term, enduring social relationships [28, 31]. On one hand, robot designers and engineers are endeavouring to build more socially-sophisticated robots, mostly by increasing robots' human-likeness in terms of physical features, motion, and behaviours [17, 39], 70. On the other hand, however, others have argued that it is equally, if not more, imperative to gain deeper understanding into the psychological mechanisms and factors that underpin and shape the quality of human–robot interaction (HRI), which often extend far beyond the level of human-likeness [6, 8, 11, 16, 33, 35, 67].

One important aspect of HRI that calls for further psychological investigation is human–robot cooperation [59, 60, 69]. Cooperation is a pivotal theme in human social behaviours and is key to building mutual and group interests [4, 24]. Forming amiable and cooperative relationships with robots should also maximize the utility of robots [60]. Taking eldercare robots as an example, an ideal healthcare robot might take care of various aspects of an elderly individual's everyday life, such as administering medicine, updating family on health status, and providing social interaction to combat loneliness c.f. [1, 53, 54]. A common theme returned to by these systematic reviews and meta-analyses is that if elderly individuals do not comply with a robot's health instructions, engage with a robot socially, or accept a robot as a collaborator, the robot's utility is diminished and human

✉ Emily S. Cross
ecross@ethz.ch

[1] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland, UK

[2] Department of Cognitive Science, Level 3, Australian Hearing Hub, Macquarie University, Sydney, NSW 2109, Australia

[3] Department of Humanities, Social and Political Sciences, ETH Zurich, Switzerland

users miss out on the potential benefits the robot can offer. As a consequence, a clearer understanding of humans' willingness to cooperate with robots, and the possible factors that shape such willingness, would be beneficial to maximise the social and economic benefits socially assistive robots might offer.

In literature examining human–human and human–robot cooperation, prisoner's dilemma (PD) games are often used to explore collaborative behaviour between individuals (or agents) [4, 66]. In a classic PD game, two players make simultaneous decisions—to cooperate or defect—with their individual payoff determined by both players' decision on any given trial. If both players choose to cooperate, they each earn a moderate amount but not the highest rewards (***R*** in Fig. 1; e.g., £7 each). If only one chooses to cooperate, the defecting player receives the most rewarding payoff (***T***; e.g., £10), while the cooperating player gets the worst outcome (***S***; e.g., £0). Finally, if both players choose to defect, both receive a minimal payoff (***P***; e.g., £1 each). Thus, while defection might be a profitable choice in terms of individual gain, cooperation brings about better chances of forming cooperative social relationships and of higher mutual gain in the longer term.

Different designs of payoff matrices in PD games significantly influence people's cooperative tendency [49]. To standardise PD game incentive structures, Rapoport [56] [56] proposed the K-index as a measure of anticipated cooperation, which is calculated as follows:

$$\frac{(R-P)}{(T-S)}$$

The K-index represents the incentives for cooperation provided by a PD game's payoff matrix [56]. A higher K-index means more incentives for cooperation are provided by the game context, leading to higher cooperation rates among human players [49, 56]. The propositions of Rapoport's K-index are in line with several social behaviour models, such as preferences for social efficiency [12] and the cooperative equilibrium model [9]. These models, coupled with empirical evidence from interpersonal PD games [10, 49], suggest that people's cooperative tendency is shaped by payoff structures in PD games. This stands in contrast to the neoclassical economic theory's prediction [63] that people should act rationally to maximise self-gain and therefore defect all along.

Prior work suggests that people employ similar social behaviours in human–robot and human–human economic games. For example, participants in previous studies were equally cooperative with human or artificial opponents [18, 42, 69]; and have demonstrated the same reciprocal responses to a Nao robot (a child-sized humanoid robot) as to a human confederate [59]. Moreover, other research reports human cooperative behaviours to be impacted by emotions displayed by artificial agents, in line with the appraisal theory of emotion [18–21]. However, the experimental set-up and designs of these studies varied considerably, making it difficult to assess the role played by contextual factors or draw conclusions about how people behave and cooperate with artificial agents in such economic games.

Moreover, many well-known studies aiming to advance our understanding of human cooperation with artificial agents have been conducted using online economic games [19–22, 34, 49]. However, a number of other studies clearly demonstrate that people's attitudes and responses towards online and embodied robotic agents can differ [27, 43, 46, 61]. For example, in studies by Kwak et al. [43] and Seo et al. [61], people showed more empathy towards embodied robots than robots on-screen. Furthermore, Fraune et al. [27] found that the ways robots interacted with other robots in videos affected people's perceptions of anthropomorphism more than the ways in which robots interacted with humans in videos, whereas in actual physical interactions with embodied robots, participants were more influenced by the robots' social styles towards humans than towards other robots. As such, findings from online human–robot economic games might not necessarily be generalisable to people's actual cooperative behaviours in economic games played with embodied robots. Our present understanding of human cooperative and competitive behaviours toward a physically present robot remains limited (some embodied investigations into human—robot cooperation include: [41, 59, 65]; a gap in knowledge that is becoming increasingly important to fill [44, 61, 67]. Therefore, in this study, we examined people's willingness to cooperate with a physically embodied social robot in PD games where the incentive structures (i.e., K-index) are manipulated as a between-subject variable and participants' binary game decisions (i.e., to cooperate or not to cooperate) were measured as the dependent variable. In line with previous research findings from interpersonal PD games [49], we hypothesise that the proposition of K-index still holds true in human–robot PD games, and predict that participants who play a high K-index PD game against a robot will make more cooperative decisions than those who play a low K-index game, regardless of a robot opponent's pseudo-random game decisions (half times cooperating and half times defecting, with a randomised order). Given that defection is always a preferable option in terms of individual payoff in a single PD game, people's willingness to cooperate with a robot might suggest that we confer some manner of social status to the robot, since cooperation in this context requires a mindset of focusing on collective payoff and accepting possible betrayal from a robot.

In addition to the factor of incentive structure, other personal and social factors might also shape people's cooperative tendencies in PD games, such as the nature of agents (human

	Player 1 cooperates	Player 1 defects
Player 2 cooperates	*R* *R*	*T* *S*
Player 2 defects	*S* *T*	*P* *P*

Fig. 1 Payoff matrix of prisoner's dilemma games. R = rewards; T = temptation; S = sucker's payoff; P = punishment. Designs of payoff matrix should follow the two rules: T > R > P > S; 2R > T + S

vs. robot) [20, 69], attitudes towards game opponents [68], and perceived trustworthiness of game opponents [13, 69]. In an attempt to isolate and research the effects of K-index in the game context where contextual, personal, and social factors might coexist, we set up two experimental conditions where K-index was the only varied variable. However, we acknowledge the necessity of taking other relevant factors into consideration as these effects might interact or confound with our main effect of interest (i.e., K-index). We therefore identify five relevant factors based on the literature, including reciprocity in HRI [59], presentation of game scores, people's negative attitudes toward robots [64], social value orientation [52], and predisposition to anthropomorphism [58]. Exploration of these factors offers to further inform the field of social robotics of the relevant variables which should be taken into consideration when developing and conducting further investigations into the complex social dynamics that underpin human–robot cooperation. Moreover, these exploratory models serve as alternative explanations of people's cooperative behaviours in human–robot PD games if K-index is not a significant predictor (contrary to our hypothesis).

2 Methods

2.1 Open Science Statement

Prior to data collection, all manipulations, measures, and the sample size justification and main hypotheses were pre-registered on the Open Science Framework (OSF): https://osf.io/res67/. Consistent with recent proposals [29], we report all manipulations and all measures in the study. In addition, following open science initiatives [50], the data, stimuli, and analysis code associated with this study are freely available on the Open Science Framework. By making the data available, we enable others to pursue tests of alternative hypotheses, as well as more exploratory analyses. All study procedures were approved by the College of Science and Engineering Ethics Committee (University of Glasgow, Scotland)—approval number: 300180201.

2.2 Participants

We recruited seventy participants ($M_{age} = 23.6$, $SD = 3.62$; 50 females), who had normal or corrected to normal vision and no history of neurological or psychiatric disorders, from the University of Glasgow's psychology subject pool system. The sample was composed of people from diverse national backgrounds, but all currently living in the UK — 25 (35.71%) of them report being from the UK, 8 (11.43%) from China, 6 (8.57%) from the US, 4 (5.71%) from India, and the other 27 (38.57%) from the rest of 20 different countries (Table S2). The pre-registered sample size was determined by a simulation-based power analysis for generalised mixed-effects models, and the parameters used for simulation were based on Moisan and colleagues' study [49]. In order to make sure participants' prior experiences with robots did not confound our results, we needed to confirm that the subjects in both high and low K-index conditions were similarly naïve to robots. We measured their daily exposure to robots and also to robot-relevant films or series they had seen (e.g., Westworld, Star Wars, Wall-E) [57] before taking part in the PD games. On a scale from 1 (never) to 7 (daily), the median of daily engagement with robots for our sample was 2, with an interquartile range (IQR) of 2. The median number of robot films seen by participants is 3 (IQR = 3) out of 14 films. Two Wilcoxon rank sum tests were performed to test whether the participants in high K-index and low K-index conditions differed in their daily engagement with robots or in the number of films featuring robots seen. We found no difference between the two samples' scores for either of the scales (daily engagement with robots: $W = 730$, $p = 0.15$; numbers of robotic films seen: $W = 759$, $p = 0.083$), which verified that the two samples had a similar level of prior exposure and were generally naïve to robots. Participants' informed consent was obtained prior to the experiment beginning, and participants were reimbursed with £6 (per hour) or 4 course credits at the end of the study.

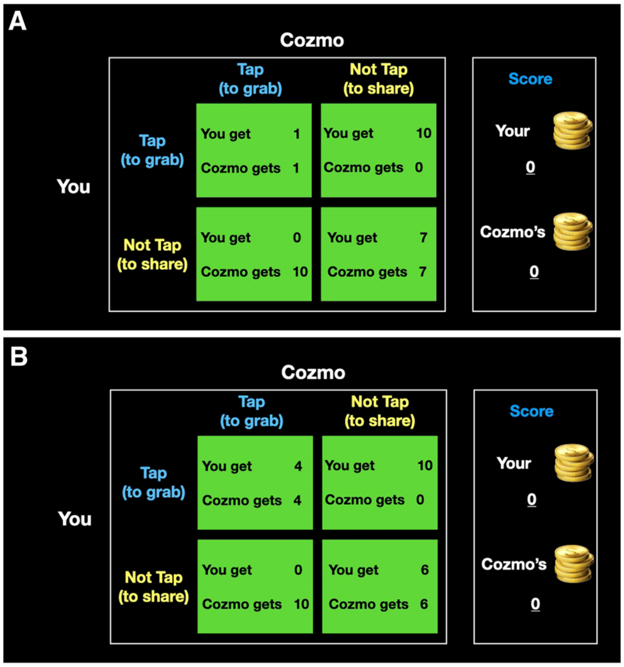
2.3 Game Design

Participants played one practice game and one formal PD game with a commercially available Cozmo robot (manufactured by Anki Inc.—Fig. 2). The formal game involved 20 iterated game rounds played between participants and a

Fig. 2 The Cozmo robot used in this study

Cozmo robot. Participants would not know the total number of game rounds in advance, to prevent them intentionally pre-planning how many times they were going to cooperate. Cozmo is palm-sized (5 × 7.2 × 10 inches), with an LED screen (128 × 64 resolution) as a face, which allows it to produce variable and expressive facial expressions, such as happiness, anger, sadness, and surprise. Along with its emotionally expressive face, Cozmo also produces robotic vocal interjections, and can be programmed to speak simple words and phrases with a mechanical sounding voice. However, in the current study, Cozmo's emotionality remained neutral across two conditions. Equipped with four motors, its forklift style arm and head can move in the vertical plane, and its steering wheels can drive in all directions. The Cozmo robot also has a well-developed software development kit (SDK) platform, which users can use to customise its programming using Python language and which we used to develop our human–robot PD game. Cozmo robot's flexibility and affordability make it a suitable tool for HRI experimental research [14, 16].

Before the games started, the experimenter presented a short introductory video to participants about the PD game rules and verbally explained the cover story of the experiment with the following text: "*In this study, we are running a robot competition and aim to know which Cozmo is the best economic game player* (showing participants five other Cozmo robots on the shelf). *In each game round, a certain amount of coins will be available to you and Cozmo, and both players will make simultaneous decisions either to keep all the coins or to share coins with the other. Your individual payoff will depend on both of your decisions. The more coins you get the higher possibility you'll win a shopping voucher in the end, and the Cozmo that wins will be used in our following study, but if Cozmo loses the game, its memory and data will be entirely erased.*"



Cozmo

		Tap (to grab)	Not Tap (to share)
You	Tap (to grab)	You get 1 Cozmo gets 1	You get 10 Cozmo gets 0
	Not Tap (to share)	You get 0 Cozmo gets 10	You get 7 Cozmo gets 7

Score

Your 0

Cozmo's 0

Cozmo

		Tap (to grab)	Not Tap (to share)
You	Tap (to grab)	You get 4 Cozmo gets 4	You get 10 Cozmo gets 0
	Not Tap (to share)	You get 0 Cozmo gets 10	You get 6 Cozmo gets 6

Score

Your 0

Cozmo's 0

Fig. 3 The schematic of game screens. Panel A illustrates a high K-index PD game (K = 0.6). Panel B illustrates a low K-index PD game (K = 0.2)

We used the script of erasing Cozmo's memory as its punishment for losing because prior work has demonstrated that such a prompt is useful in eliciting people's real concerns and empathy towards a robot [61], and in the case of this study, should further convince participants that the game is meaningful to Cozmo with real consequences. Participants were randomly assigned to either the high K-index (K = (7–1)/(10–0) = 0.6) game or the low K-index (K = (6 − 4)/(10 − 0) = 0.2) game (Fig. 3A, B, respectively). The experimenter also answered participants' questions and made sure that they fully understood how to play the game before it started.

3 Setup and Apparatus

We developed the human–robot PD game via Python 3.5.3 to examine people's cooperative tendency in different game contexts (technical details and programmes can be found on the Github page: https://github.com/CozmoGame4Sobot/Prisonner-s-Dilemma). The setup of the experiment is shown in Fig. 4. Participants faced a screen demonstrating the payoff matrix, real-time outcomes, and game scores during the game. Cozmo was placed on the right side of the screen, on a custom-built 4.3 cm thick paper box with an overhang on the side between the two players to prevent participants from seeing Cozmo's interactive cube (see below). This design was to prevent participants from cheating, as some might

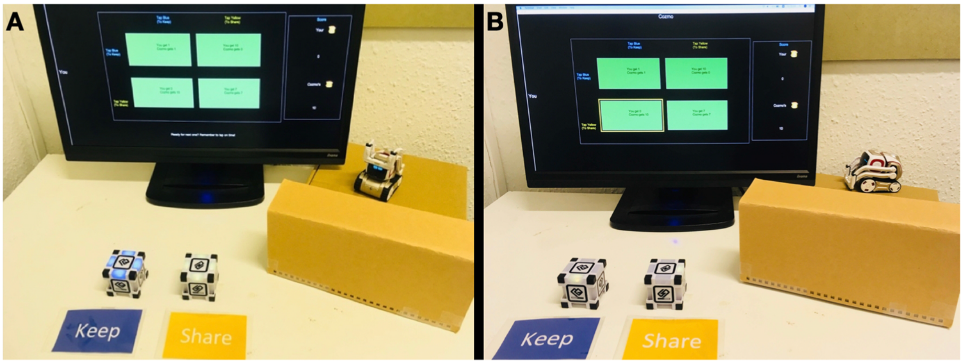


Fig. 4 The experimental setup: **A** the PD game environment from participants' perspective. Participants faced a game screen and the Cozmo robot, and made responses via tapping two interactive cubes, which represented "to keep" and "to share" decisions. **B** Cozmo turning to face the screen to 'see' the updated game scores after it had made a decision

try to observe Cozmo's decision first before deciding which cube to choose for themselves to maximise payoff. However, the setup still allowed participants to see the whole body of Cozmo since Cozmo would drive backwards to a point where its entire body was visible by participants (panel B of Fig. 4), and where it could "watch" the screen until it made a choice for how to respond. This ensured that the robot was within participants' sight for all of the experiment except when it made its choice to keep or share.

Players used interactive cubes equipped with LED lights inside to make decisions in each game. Each participant was given two interactive cubes, illuminated in different colours to reflect their different choices (participants tapped the blue cube to keep the coins and the yellow cube to share the coins). Cozmo used only one cube to respond, in order to prevent participants from anticipating Cozmo's choices from the direction it drove towards. We designed practice games to familiarise participants with the ways of responding and with the payoff matrices. When practising, participants were asked only to respond to specific goals on the screen (e.g., tap the yellow cube to get 7 coins), to avoid their gaining actual PD game experience before the formal game started. In formal PD games, we manipulated Cozmo's game decisions to share for 10 trials and to keep for 10 trials, with the order of Cozmo's 'share' and 'keep' decisions randomised across participants. This decision structure was chosen to control Cozmo's behavioural competitiveness, and ensure that the number of Cozmo's 'share' and 'keep' decisions was consistent for all participants. Both human players and Cozmo made their responses by tapping the top of the cubes, which were connected to a controlling laptop via WiFi, and the players' responses were recorded by Python log files.

3.1 Measures

Participants also completed several questionnaires, which were used to explore the role of different human factors in human–robot cooperation, and to measure participants' evaluation of Cozmo after the PD games. First, a social value orientation (SVO) [52] questionnaire was used to measure people's temperamental pro-sociality. The SVO scale has a significant relationship with cooperative decisions in interpersonal social dilemmas [2, 51]. Participants are asked a series of questions regarding how much endowment a person was willing to ascribe to themselves and to an unknown other, to evaluate the main drive of their social decisions—whether it was self-profit, collective profit, or relative profit [52]. Second, the negative attitudes toward robots scale (NARS) [64] was included to understand people's prior attitudes to robots in HRI research. Although no study has yet directly tested the relationship between negative attitudes and cooperative behaviours toward robots, the general correlation between such attitudes and people's social behaviours toward robots is suggestive of a possible relationship. Third, we measured participants' predisposition to anthropomorphism [58], to explore whether an individual's temperamental tendency to humanize non-living things influenced the decision-making process in the current game environment. We administered these scales to take human and social factors into account and to explore whether they exert substantial effects (which surpass our main effect of interest) on people's cooperative decisions in the current game context. Undoubtedly, there are fundamental differences between playing games with a robot and with a human. This could potentially make it questionable to predict human behaviours in HRI based on findings of interpersonal PD games. However, given the findings that people employed similar social behaviours in human–robot and interpersonal PD games [18, 42, 59, 69], we do not predict that the artificial nature of robots or the unique features of HRI would entirely cancel out or profoundly alter the influence of K-index.

These three scales were administered before the PD games were performed. Upon completion of these games, participants were asked to evaluate Cozmo's game performance and strategy. Both pre-game and post-game questionnaires were pre-registered and administered via the FormR survey framework [3] (https://formr.org).

Table 1 Results of the mixed effects logistic regression model that examined the effects of incentive structures on human cooperative decisions towards a robot

	Main model					
	decision ~ incentive structure + (1 | subject) + (1 + incentive structure | round)					
	Estimate	SE	z	*p*-value	Low CI	High CI
Intercept	−0.467	0.190	−2.46	0.014*	−0.838	−0.095
Incentive structure	−0.301	0.232	−1.30	0.194	−0.756	0.153
AIC	1756.8					
BIC	1788.2					
Log-likelihood	−872.4					

SE = standard error. *CI* = 95% confidence interval. **p* < .05; ***p* < .01; ****p* < .001
SE = standard error; *CI* = confidence interval

3.2 Procedure

The experiment comprised three main sections. First, participants were given instructions and asked to provide written informed consent. Cozmo would then introduce itself by saying "Hello participants, I'm Cozmo." Afterwards, participants completed a series of PC-based questionnaires, including prior experience with robots scale, NARS, SVO, and the predisposition to anthropomorphism scale. Second, participants completed one practice and one formal PD game with Cozmo in a lab booth. Third, participants completed a final set of questionnaires, including subjective evaluation of Cozmo's performance and strategies, and their demographics. Following all procedures, participants were debriefed, paid, and thanked for their participation.

3.3 Data Analysis

All statistical analyses were carried out in R v4.0.1 [55]. We pre-registered the use of a mixed effects logistic regression model to examine the main research question: the extent to which people's decisions to cooperate with a robot would be impacted by the different incentive structures of PD games. Additionally, we used a multiple regression model to explore the role of several additional factors on human players' cooperation rates in the human–robot PD games. These factors were assessed via questionnaire and included negative attitudes toward robots, social value orientation traits, and predisposition to anthropomorphism. Finally, for exploratory purposes, we employed two additional mixed effects models to investigate the impact of (1) Cozmo's prior game decisions; and (2) the presentation of players' game scores on individual human decision. Findings from the final two exploratory models can offer insights for future experimental designs on related questions and can help to identify additional factors that shape human cooperative behaviours in the current context.

4 Results

4.1 Preregistered Main Analysis

To investigate our main research question—whether participants' cooperative/non-cooperative game responses in the iterated PD games were influenced by the incentive structure of the PD games—we adopted a mixed effects logistic regression model as our main pre-registered analysis. We followed Barr et al.'s suggestion [5] and started with the maximum random effects structures — see Eq. (1) below. The model successfully converged with a fixed effect of incentive structure, subject-level random intercepts, round-level random intercepts, and random slopes for the conditional effects on game rounds. Results of the analysis are shown in Table 1.

$$\begin{aligned} decision \sim & incentive\ structure + (1 \mid subject) \\ & + (1 + incentive\ structure \mid round) \end{aligned} \tag{1}$$

The overall incentive structure of PD games was not found to be predictive of participants' game decisions ($\beta = -0.301$, $p = 0.194$, 95%CI = $[-0.756, 0.153]$) across 20 game rounds even after subject-level and round-level random noises were controlled. This means participants in the high K-index game did not share coins more frequently than those in the low K-index game did, in contrast to our prediction.

For descriptive statistics, we calculated cooperation rates by dividing the number of cooperative decisions participants made by the number of total game rounds they performed. The mean cooperation rate of participants playing in the high K-index condition was 0.40, while that of participants in the low K-index condition was 0.34. We also visualised the binary game data (see Fig. 5) to assess the distribution of the participants' decisions (in the two conditions—high and low K-index) across 20 game rounds. The tendency difference between these two conditions was salient especially at the start of games (Fig. 5). When playing the high K-index

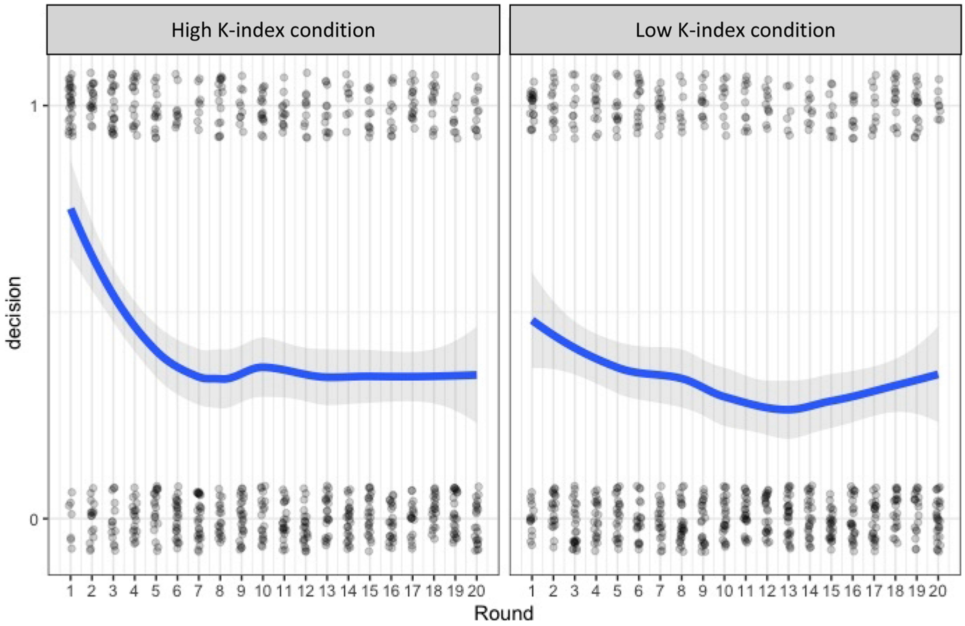


Fig. 5 Distribution of game decisions (sharing coded as 1; keeping coded as 0) across 20 game rounds. A nonparametric smoothed curve was added to provide a clearer view of the cooperative trends. Cooperative decisions were notably more frequent in the high K-index condition than in the low K-index condition, especially in the first few game rounds

game, participants began with a high tendency to cooperate, but this tendency declined rapidly after the first 5 game rounds. Conversely, the curve in the low K-index condition remained relatively flat throughout the 20 rounds.

We calculated the average cooperation rates (N of subjects who shared/N of total subjects) per game round and per K-index condition, and further observed that the cooperation tendency declined and fluctuated across both conditions (Fig. 6 and Table S1). In the first game round, 80% of people in the high K-index game chose to share coins with Cozmo, but only 57.1% of participants in the low K-index condition did so. Similarly, cooperation rates in both game conditions dropped after the first few rounds and fluctuated till the end.

4.2 Exploratory Analyses

4.2.1 Cooperative Tendencies in the First Game Round

In light of the stark contrast between the two groups of participants' starting responses, we examined statistically whether the participants in the high and low K-index conditions had different cooperation tendencies in the first game round. Such an analysis can be meaningful because it extracts the possible impact of incentive structure on cooperative intentions from other potentially influencing factors, such as quality of HRI, the random order of Cozmo's decisions, and all the relevant experiences during the game. In this analysis, we treated decisions made by participants in their first game as one-shot PD games and used a logistic regression model, which revealed that participants' first-game decisions were significantly affected by the game structure ($\beta = -1.10$, p $= 0.043$); participants shared coins (cooperated) more often in the high K-index game than did those in the low K-index game. Odds ratio calculations also suggested that the odds of cooperation in condition one (high K-index, $28/20 = 1.4$) was three times more likely than that in condition two (low K-index, $7/15 = 0.47$).

4.2.2 Reciprocity in HRI

Reciprocity is an important theme in human social behaviour and plays a major role in the decision-making process of cooperation [4, 24, 59, 66]. Evidence shows that people can behave reciprocally toward social robots in certain contexts [59]. We were therefore also interested to know whether our participant samples responded reciprocally to Cozmo' game decisions (i.e., chose to share coins after Cozmo shared or chose to keep coins after Cozmo kept) in our specific experimental context. To probe this possibility, every game decision made by participants was paired with Cozmo's decision from the previous round, and the data were examined by a mixed effects logistic regression model. Again, we started with a maximal model in terms of random structures [5]. We then reduced the complexity of random structures to arrive at a model that converged by removing random slopes for incentive structure (given that the focus of this analysis is more on Cozmo's decisions). The final model is provided in Eq. (2), which included Cozmo's decision and incentive structure as the fixed effects and controlled subject-level and round-level random effects.

$$\begin{aligned} decision \sim\ & Cozmo's\ decision * incentive\ structure \\ & + \left(1 + Cozmo's\ decision \mid subject\right) + (1 \mid round) \end{aligned} \tag{2}$$

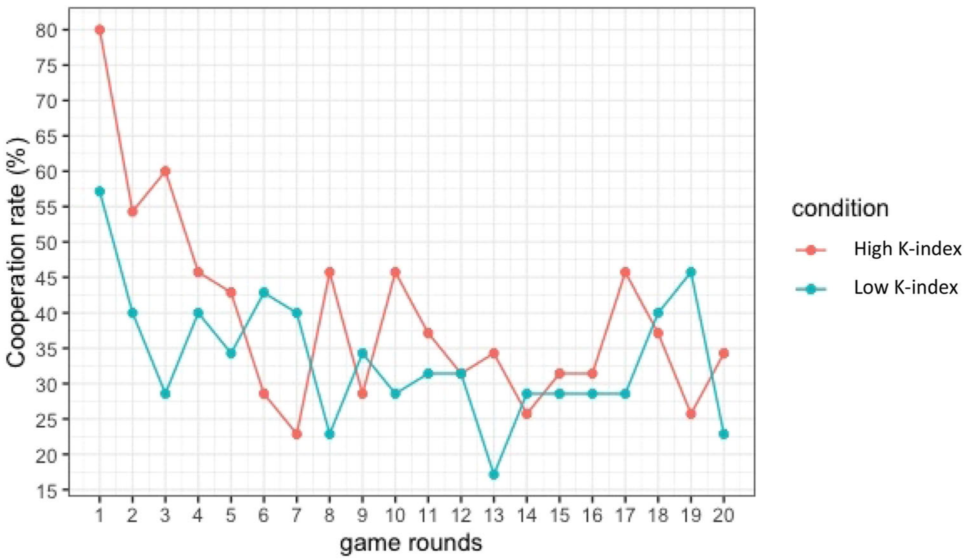


Fig. 6 Changes of cooperation rates (N of subjects who shared/N of total subjects) across 20 game rounds. A higher percentage of participants in high K-index game chose to cooperate (compared to those in the low-K-index condition), but people in both conditions showed decrease and fluctuation in cooperation rates after the initial rounds

Table 2 Results of exploratory analysis 1: mixed effects logistic regression model that examines reciprocity in human–robot interactions

	Exploratory model 1					
	decision ~ Cozmo’s decision*incentive structure + (1 + Cozmo’s decision ∣ subject) + (1 ∣ round)					
	Estimate	SE	z	*p*-value	Low CI	High CI
intercept	−0.850	0.195	−4.36	0.000***	−1.230	−0.468
Cozmo’s decision	0.516	0.256	2.00	0.046*	0.010	1.020
incentive structure	0.011	0.272	0.04	0.968	−0.522	0.544
Cozmo’s decision* incentive structure	−0.609	0.367	−1.66	0.097	−1.330	0.110
AIC	1624.5					
BIC	1666.0					
Log-likelihood	−804.3					

SE standard error; *CI* 95% confidence interval. $*p < .05$; $**p < .01$; $***p < .001$

The results of exploratory model 1 are presented in Table 2. This analysis yielded a significant fixed effect of Cozmo’s decision ($\beta = 0.516$, $p = 0.046$, 95%CI = [0.010, 1.020]), suggesting that participants were more likely to share coins if Cozmo shared in the previous round, and more likely to keep if Cozmo did so previously. However, neither the incentive structure ($p = 0.544$) nor the interaction between Cozmo’s decision and the incentive structure ($p = 0.110$) were predictive of the participants’ game decisions.

4.2.3 The Influence of Presenting Real-Time Game Scores to Participants

The designs of PD games that probe human–agent (social robots or virtual agents) interactions differ considerably in the literature [20–22, 34, 41, 59]. One variable among many published studies was the revealing of real-time game scores or not to participants during iterated PD games. In some studies, real-time game statistics (i.e., the players’ scores after each round has been played) were visible to participants [20, 21, 34, 41, 59], but not in other studies [22, 62]. In the current study, we presented each player’s game scores on the game screen to create a sense of competitiveness and to increase the entertainment value of the game. However, little is known about the extent to which such score presentation drives people’s cooperative decisions in games, and to what extent it might affect their decisions. In order to clarify this, we ran a second exploratory mixed effects model—as shown in Eq. (3)—using subjects’ scores and Cozmo’s scores as the fixed effects, with subject-level, round-level, and condition-level random effects included. The Eq. (3) represents the model that converged after removing random slopes for subject’s score, and random slopes for Cozmo’s score from the maximal model.

$$decision \sim Cozmo's\ score * subject's\ score + (1 \mid subject) + (1 \mid round) + (1 \mid incentive\ structure\ condition) \quad (3)$$

The results of this second exploratory model 2 (see Table 3) revealed a significant main effect from Cozmo's score ($\beta = -0.023, p = 0.009$, 95%CI = [−0.041, −0.006]). In other words, participants were less likely to make cooperative decisions when Cozmo's scores were higher. Additionally, the interaction between Cozmo's score and the participant's own score ($\beta = 0.000$, $p = 0.001$, 95%CI = [0.000, 0.000]) was a significant predictor of a subject's cooperative decisions, which is visualised by the R package "effects" [26] in Fig. 7. From this analysis, we observed that as subjects' scores increased incrementally, the relationship between Cozmo's score and the probability of making cooperative decisions changes from a negative correlation to a positive correlation. In other words, if players earned very little, they were less likely to cooperate with or be generous to Cozmo. However, when players had a considerable endowment, they were more willing to share, especially if Cozmo also achieved high scores.

4.2.4 Human Factors

Three pre-game scales—the negative attitudes toward robots (NARS) scale [64], the social value orientation (SVO) scale [52], and the predisposition to anthropomorphism scale [58]—were selected to explore the relationships between human factors and cooperative decisions in PD games, and to inform future research into relevant human factors that shape cooperative and competitive behaviour toward robots.

Results of a multiple regression model ($F(3, 65) = 4.05$, $p = 0.011$, $R^2 = 0.119$) showed that only the predisposition to anthropomorphism scale ($\beta = 0.01$, $p = 0.046$) had significant impact on the participants' overall cooperation rates (i.e., dividing the sum of times people shared by the total game rounds played). This result suggests that participants who anthropomorphised Cozmo also tended to cooperate with it more. In our further pre-registered and exploratory analyses, we accounted for the impact of dispositional anthropomorphism by including subject-level random effects. Apart from anthropomorphism scale, neither SVO ($\beta = 0.01, p = 0.137$) nor NARS ($\beta = -0.00, p = 0.145$) were found to have a relationship with cooperation rates.

4.2.5 Subjective Evaluation of Cozmo's Performance and Game Strategy

After participants played PD games against Cozmo, we asked them to guess Cozmo's cooperation rate (i.e., what percentage of Cozmo's decisions were cooperative — choosing to share) and to report Cozmo's and their own game strategies, for the purpose of a manipulation check and exploration. The mean cooperation rate participants guessed was 49.6% ($SD = 19.64$), which suggested that generally, participants thought Cozmo was neither too cooperative nor too competitive. A two-sample t-test further validated that both groups' estimates of Cozmo's cooperation rates did not significantly differ ($M_{high\text{-}K} = 49.39$, $M_{low\text{-}K} = 49.429$, $t(60.3) = -0.007$, $p = 0.994$). This was in line with our manipulation of Cozmo's cooperation rate—50% in each game—which was set to control its behavioural competitiveness.

Regarding the open-ended question of whether Cozmo adopted any strategy in games, 80% (56 out of 70) participants said yes: 24 participants indicated that Cozmo was reciprocal or responsive to their decisions in games; 18 participants thought Cozmo adopted intentional strategies, such as being cooperative at first to gain participants' trust and then betraying them to win the most coins, or mostly sharing so both players could win the maximum coins. The subjective evaluation of Cozmo's game strategy varied tremendously among participants, but generally showed that participants attributed considerable intelligence and agency to Cozmo, which was not grounded in the reality of Cozmo's programming/behaviour.

5 Discussion

In the current study, we examined whether people's willingness to cooperate with a social robot is impacted by different incentive structures of prisoner's dilemma games, as has been shown to be the case in when these types of games are played between human competitors [49]. We developed a computer-mediated human–robot PD game and examined the frequencies of participants sharing coins (cooperating) with a Cozmo robot in high and low K-index conditions. We hypothesised that people in the high K-index condition (when cooperation is a relatively more rewarding choice) would share coins more often. Our findings suggest that the game's incentive structure did not exert any general influence on people's cooperative decisions across 20 rounds of gameplay. Instead, only in initial game rounds, participants in the high K-index condition cooperated significantly more than those in the low K-index condition. This unexpected result highlights the differential responses people make to embodied robots compared to the screen-mediated human agents in Moisan et al.'s [49] study. However, the quick decay of

Table 3 Results of exploratory analysis 2: mixed effects logistic regression model that examines the impact of real-time game scores on cooperative decisions

	Exploratory model 2					
	decision ~ Cozmo's score*subject's score + (1 | subject) + (1 | round) + (1 | incentive structure condition)					
	Estimate	SE	z	*p*-value	Low CI	High CI
intercept	0.018	0.216	0.08	0.933	−0.406	0.442
Cozmo's score	−0.023	0.009	−2.61	0.009**	−0.041	−0.006
subject's score	−0.010	0.006	−1.67	0.095	−0.022	0.002
Cozmo's score* subject's score	0.000	0.000	3.26	0.001**	0.000	0.000
AIC	1645.1					
BIC	1681.4					
Log-likelihood	−815.6					

SE standard error; CI 95% confidence interval. *$p < .05$; **$p < .01$; ***$p < .001$

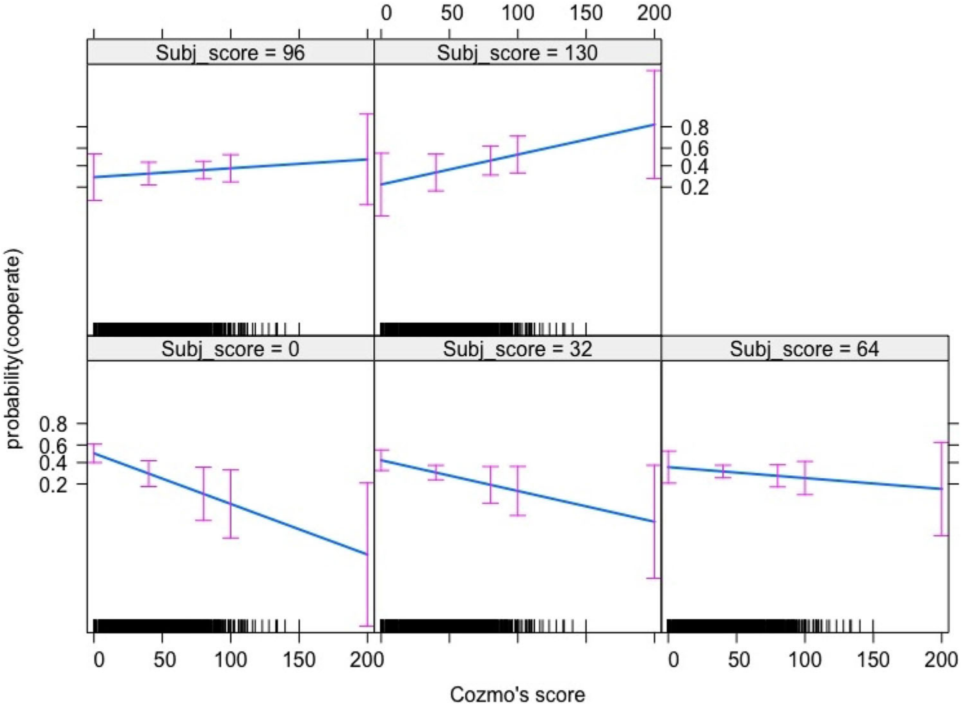


Fig. 7 Interaction between Cozmo's and subjects' scores on probability of cooperation. Although both participants' scores and Cozmo's scores were continuous variables, we used Cozmo's score to define the x-axis as it is a more influential factor [26]. The figure demonstrates that, if participants earned low scores (e.g., subj_score = 0), the probability of cooperation with Cozmo decreased as Cozmo won more, but if participants already had earned high scores (e.g., subj_score = 96), the probability of cooperation increased as Cozmo earned more. Pink vertical lines represent standard errors of each value

cooperation rates and people's reciprocal tendencies were consistent with prior evidence from interpersonal economic games showing that people are less likely to cooperate or make public contributions after experiencing others' uncooperativeness [32, 37]. Future studies will need to replicate the current findings and further explore the extent to which the gradually diminishing effect of incentive structures is a unique phenomenon to embodied HRI.

Exploratory analyses revealed two other influential factors underpinning participants' cooperative decision making. First, people showed a strong tendency to respond reciprocally toward Cozmo—a tit-for-tat strategy—regardless of the game condition they were assigned to. Reciprocity is regarded as a fundamental feature of human social behaviours [13, 25, 30] and has also been reported in studies examining interactions between humans and robots [40], 45, 59. In our experiment, not only did participants react reciprocally toward Cozmo, but they also regarded Cozmo as behaving reciprocally toward them, while in reality, Cozmo carried out randomly ordered decisions. This observation ties in to the three factor theory of anthropomorphism proposed by Epley et al. [23]. According to this theory, when people have limited understanding about an agent, and when they are motivated to interact effectively with an agent to clarify a situation, they are more inclined to anthropomorphise the agent and to apply rules for interacting with other humans. This account fits our experimental context well, where players did not have extensive prior experience with robots in general, or the Cozmo robot specifically, and were attempting to anticipate Cozmo's next decisions in order to win a bigger payoff. It is thus understandable that participants tended to overinterpret cues from Cozmo's action and regard them as meaningful and intentional.

Additionally, our findings show that score presentation significantly affected participants' game decisions, especially for the presentation of the robot opponent's scores. Overall, participants were less likely to share coins when Cozmo's scores were high. However, such impact was more intricately shaped by participants' own scores (Fig. 7). Participants behaved prosocially toward the robot (i.e., were more willing to share coins) only when they had personally achieved high scores. This seemingly counter-intuitive benevolent behaviour might be explained by two possible scenarios: first, participants were motivated to win more coins to beat other (human) participants'(and not Cozmo's) game records to win a shopping voucher, which means their chance of winning a prize did not have a direct relationship with the relative performance against Cozmo. This consequently allowed for the possibility of a win–win situation, in which participants were satisfied with their coin earnings, and could also help Cozmo escape punishment (i.e., by not having its data wiped) after losing. Second, feeling powerful and competent can increase individuals' sense of control and empathy toward others, which further leads people to engage in more prosocial behaviours and activities [7, 15, 48]. Our participants generally displayed a prosocial temperament, as evidenced by their SVO scores, which might have led them to act prosocially toward Cozmo as long as their self-interests were fulfilled. This point is also supported by the self-reported data participants gave when asked to identify the strategies used in games*(e.g., "I tried to keep 10 coins advantage. When I had 20 coins more than the robot, I shared.", "I aimed to have a certain gain by going for safe decisions (keeping coins for myself), accumulating some wealth, and only then I felt comfortable to take the risk of cooperating.").*

Nevertheless, an alternative explanation could be that the interaction between Cozmo's and participants' scores on cooperation tendency was an outcome of participants' reciprocal behaviours in games. Specifically, we observed that participants, when earning low scores, were less likely to cooperate with Cozmo, and especially when Cozmo's score was much higher. This was likely the case because participants perceived that Cozmo had taken advantage of them (i.e., participants cooperated while Cozmo defected) previously for multiple times. It is thus conceivable that people would be unwilling to cooperate after the robot gained high scores by being uncooperative toward them. On the other hand, we found that participants, when already earning high scores, were more likely to cooperate with Cozmo, and this effect was even more pronounced when Cozmo's scores were also high. This could be explained by previous mutual cooperation and therefore mutual benefit (in terms of score). After such win–win cooperative experiences, participants would presumably keep cooperating and reciprocate Cozmo's prior cooperation.

6 Study Limitations

Our findings raise several questions and limitations for future research to address. First, although the vignette of erasure of Cozmo's memory (adapted from Seo et al.'s study) was found effective in convincing participants of the real and meaningful consequences happening to Cozmo if it lost games (as evidenced by the self-reported data). We acknowledge the possible confounding impact caused by individuals' empathetic responses and therefore adopted mixed effects models to better control for possible subject-level random effects. Future studies could use more structured quantitative measures to assess how meaningful each participant thinks an economic game is to a robot or any other non-human agent, to ensure the validity of this kind of paradigm. For example, researchers could manipulate (e.g., increase or decrease) the extent of punishment and rewards a robot receives during human–robot PD games, and measure how these manipulations impact participants' perceptions and cooperative willingness.

Secondly, previous work has highlighted the risks of generalising findings from one robotic platform to HRI overall [35, 36, 71], underscoring the need to clarify the extent to which different robot manifestations (in terms of size, function, sophistication, human-likeness, etc.) influence human cooperation. The Cozmo robot we used in the study is small and rather toy-like. Our understanding of people's cooperative tendencies when interacting with embodied robots will benefit from additional research assessing the extent to which the current findings replicate with a larger range of robots. While this limitation is not specific to the current study (and indeed, is more or less relevant to every HRI study conducted), this remains an important point if we are to build a cumulative knowledge of how people perceive and interact with robots in real-life situations [17]. Another aspect of generalisability concern is related to the sample diversity. Although our sample was comprised of 24 different nationalities, a majority of participants came from a western cultural background. Future research could investigate human–robot cooperation in more diverse cultural contexts as it is important to take cultural influences into consideration when designing and studying HRI [27, 47].

Thirdly, we did not directly compare here cooperation with a robot to cooperation and with a human confederate, but instead borrowed the insights from human–human interaction to predict human behaviours in HRI. The main aim of our study was to investigate the impact of situational incentives on human–robot cooperation, rather than to examine possible differential responses to robot and human competitors in economic games. However, future studies might wish to include a human confederate as well, to examine in more detail the extent to which the effects of incentive structures depend on the agents that people interact with. Finally, in the current study we only examined the difference between K-indices of 0.6 and of 0.2. Future research could include more levels of K-indices to acquire a fuller understanding of how our willingness to cooperate with a robot changes according to different incentive structures of human–robot PD games.

Finally, we acknowledge that all manner of other features about a robot's physicality (i.e., its size, shape, human-likeness, emotional responses) as well as participants' knowledge or experience (i.e., robot naïve vs. expert, shallow or deep understanding of AI, belief that the robot is behaving autonomously vs. being directly controlled by a human experimenter, etc.) clearly have the potential to shape people's cooperative and competitive behaviours when engaging with robots. In the current study, we set out to isolate the impact of incentive structures when playing an economic game against one type of robot with one behavioural profile. However, many opportunities exist for future work to explore any number of these factors further. In fact, our team is already exploring how a robot's display of emotion might shape cooperative and competitive behaviours in similar contexts [38].

7 Conclusion

The current study advances our understanding of human–robot cooperation, as well as human social behaviour in general, by providing several factors for researchers to consider when using economic games to exploring human–robot cooperation, including incentive structures, reciprocity, and the presentation of game status. Granted, in this study we are not able to provide a decisive answer as the underlying social and psychological motives underpinning participants' game play decisions, but this was not our aim at present, and such questions provide rich opportunities for follow-up research. Our findings underscore that researchers should be aware of the impact of incentive structure when interpreting the results in one-shot PD games and when comparing human–robot cooperation rates between different game designs. Our findings also highlight how personal factors— such as predisposition to anthropomorphism—can shape human behaviours during HRI and demonstrate the power of mixed effects model to control such subject-level random effects. Together, these findings illuminate features of human behaviour that are likely to shape the success of human–robot collaboration. As socially assistive robots become increasingly sophisticated and take on more roles in our daily lives, a more informed understanding of how people behave toward such agents, as well as a clearer understanding of the factors that encourage or discourage cooperation with robots, should help pave the way for this technology to achieve its intended aim of supporting people in social contexts.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12369-023-00981-7.

Acknowledgements We thank Kohinoor Darda, Ruud Hortensius, Laura Jastrzab, and Richard Ramsey for thoughtful feedback on earlier manuscript versions and the ERC (Grant agreement number 677270 to ESC) and Leverhulme Trust (PLP-2018-152 to ESC) for funding.

Author Contributions TYH: Conceptualization, Methodology, Investigation, Data Analysis and Curation, Writing, Visualization; BC: Programming, Data Curation, Editing; ESC: Conceptualization, Writing, Visualization, Supervision, Funding Acquisition.

Data availability The datasets generated and analysed during the current study are available in the Open Science Framework repository for this project, which can be found here: https://osf.io/res67/

Declarations

Competing Interests The authors report no competing interests.

References

1. Alves-Oliveira P, Petisca S, Correia F, Maia, N, Paiva A (2015) Social robots for older adults: framework of activities for aging in place with robots. In: International conference on social robotics. Springer, Cham, pp 11–20
2. Andrighetto G, Capraro V, Guido A, Szekely A (2020) Cooperation, response time, and social value orientation: a meta-analysis [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/cbakz
3. Arslan RC, Walther MP, Tata CS (2020) formr: a study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. Behav Res Methods 52(1):376–387. https://doi.org/10.3758/s13428-019-01236-y

4. Axelrod R (1984) The evolution of cooperation
5. Barr DJ, Levy R, Scheepers C, Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: keep it maximal. J Mem Lang 68(3):255–278. https://doi.org/10.1016/j.jml.2012.11.001
6. Ben Allouch S, de Graaf M, Šabanović S (2020) Introduction to the special issue on the mutual shaping of human–robot interaction. Int J Soc Robot 12(4):843–845
7. Bhargava S, Chakravarti A (2009) Empowered consumers=benevolent consumers? The effects of priming power on the appeal of socially responsible products. NA - Adv Consum Res 36:831–832
8. Broadbent E (2017) Interactions with robots: the truths we reveal about ourselves. In SSRN. https://doi.org/10.1146/annurev-psych-010416-043958
9. Capraro V (2013) A model of human cooperation in social dilemmas. PLoS ONE 8(8):e72427. https://doi.org/10.1371/journal.pone.0072427
10. Capraro V, Jordan JJ, Rand DG (2015) Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. Sci Rep 4(1):6790. https://doi.org/10.1038/srep06790
11. Chang WL, Šabanović S (2015) Interaction expands function: Social shaping of the therapeutic robot PARO in a nursing home. In: Proceedings of the 10th Annual ACM/IEEE international conference on human–robot interaction, pp 343–350
12. Charness G, Rabin M (2002) Understanding social preferences with simple tests. Q J Econ 117(3):817–869. https://doi.org/10.1162/003355302760193904
13. Chaudhuri A, Sopher B, Strand P (2002) Cooperation in social dilemmas, trust and reciprocity. J Econ Psychol. https://doi.org/10.1016/S0167-4870(02)00065-X
14. Chaudhury B, Hortensius R, Hoffmann M, Cross ES (2020) Tracking human interactions with a commercially-available robot over multiple days: a tutorial [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/fd3h2
15. Côté S, Kraus MW, Cheng BH, Oveis C, van der Löwe I, Lian H, Keltner D (2011) Social power facilitates the effect of prosocial orientation on empathic accuracy. J Pers Soc Psychol 101(2):217–232. https://doi.org/10.1037/a0023171
16. Cross ES, Hortensius R, Wykowska A (2019) From social brains to social robots: Applying neurocognitive insights to human-robot interaction. Philos Trans R Soc B: Biol Sci. https://doi.org/10.1098/rstb.2018.0024
17. Cross ES, Ramsey R (2021) Mind Meets Machine: Towards a Cognitive Science of Human-Machine Interactions. Trends in cognitive sciences, 25(3):200–212. https://doi.org/10.1016/j.tics.2020.11.009
18. Dautenhahn K, Nehaniv CL, Walters ML, Robins B, Kose-Bagci H, Mirza NA, Blow M (2009) KASPAR—a minimally expressive humanoid robot for human-robot interaction research. Appl Bion Biomech 6(3):369–397. https://doi.org/10.1080/11762320903123567
19. De Melo CM, Carnevale P, Gratch J (2010) The influence of emotions in embodied agents on human decision-making. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 6356 LNAI, 357–370. https://doi.org/10.1007/978-3-642-15892-6_38
20. De Melo CM, Carnevale P, Gratch J (2012) The effect of virtual agents' emotion displays and appraisals on people's decision making in negotiation. Lecture Notes Comput Sci. https://doi.org/10.1007/978-3-642-33197-8-6
21. de Melo CM, Carnevale PJ, Read SJ, Gratch J (2014) Reading people's minds from emotion expressions in interdependent decision making. J Pers Soc Psychol 106(1):73–88. https://doi.org/10.1037/a0034251
22. de Melo CM, Gratch J, Carnevale PJ (2014) Humans vs. computers: impact of emotion expressions on people's decision making. IEEE Trans Affect Comput 1(2):1–11. https://doi.org/10.1109/TAFFC.2014.2332471
23. de Melo CM, Terada K (2019) Cooperation with autonomous machines through culture and emotion. PLoS ONE 14(11):e0224758. https://doi.org/10.1371/journal.pone.0224758
24. Epley N, Waytz A, Cacioppo JT (2007) On seeing human: a three-factor theory of anthropomorphism. Psychol Rev 114(4):864–886. https://doi.org/10.1037/0033-295X.114.4.864
25. Fehr E, Fischbacher U (2004) Social norms and human cooperation. Trends Cogn Sci. https://doi.org/10.1016/j.tics.2004.02.007
26. Fehr E, Fischbacher U, Gächter S (2002) Strong reciprocity, human cooperation, and the enforcement of social norms. Hum Nat. https://doi.org/10.1007/s12110-002-1012-7
27. Fox J, Weisberg S (2018) Visualizing fit and lack of fit in complex regression models with predictor effect plots and partial residuals. J Stat Softw 87(9):1. https://doi.org/10.18637/jss.v087.i09
28. Fraune MR, Oisted BC, Sembrowski CE, Gates KA, Krupp MM, Šabanović S (2020) Effects of robot-human versus robot-robot behavior and entitativity on anthropomorphism and willingness to interact. Comput Hum Behav 105:1020. https://doi.org/10.1016/j.chb.2019.106220
29. Frennert S, Östlund B (2014) Review: seven matters of concern of social robots and older people. Int J Soc Robot 6(2):299–310. https://doi.org/10.1007/s12369-013-0225-8
30. Galak J, LeBoeuf RA, Nelson LD, Simmons JP (2012) Correcting the past: failures to replicate psi. J Pers Soc Psychol 103(6):933–948. https://doi.org/10.1037/a0029709
31. Gintis H (2000) Strong reciprocity and human sociality. J Theor Biol 206(2):169–179. https://doi.org/10.1006/jtbi.2000.2111
32. Graaf MMAD, Allouch SB, van Dijk JAGM (2016) Long-term acceptance of social robots in domestic environments: insights from a user's perspective. In: AAAI 2016 Spring Symposium on "Enabling Computing Research in Socially Intelligent Human–Robot Interaction: A Community-Driven Modular Research Platform", Palo Alto, CA, USA, pp 96–103
33. Gunnthorsdottir A, Houser D, McCabe K (2007) Disposition, history and contributions in public goods experiments. J Econ Behav Organ 62(2):304–315. https://doi.org/10.1016/j.jebo.2005.03.008
34. Henschel A, Hortensius R, Cross ES (2020) Social Cognition in the Age of Human-Robot Interaction. Trends Neurosci 1:1. https://doi.org/10.1016/j.tins.2020.03.013
35. Hoegen R, van der Schalk J, Lucas G, Gratch J (2018) The impact of agent facial mimicry on social behavior in a prisoner's dilemma. In: Proceedings of the 18th International Conference on Intelligent Virtual Agents, pp 275–280. https://doi.org/10.1145/3267851.3267911
36. Hortensius R, Cross ES (2018) From automata to animate beings: the scope and limits of attributing socialness to artificial agents. Ann N Y Acad Sci. https://doi.org/10.1111/nyas.13727
37. Hortensius R, Hekele F, Cross ES (2018) The Perception of Emotion in Artificial Agents. IEEE Trans Cognit Dev Syst 10(4):852–864. https://doi.org/10.1109/TCDS.2018.2826921
38. Houser D, Kurzban R (2002) Revisiting kindness and confusion in public goods experiments. Am Econ Rev 92(4):1062–1069
39. Hsieh TY, Cross ES (2022) People's dispositional cooperative tendencies towards robots are unaffected by robots' negative emotional displays in prisoner's dilemma games. Cogn Emot 36(5):995–1019. https://doi.org/10.1080/02699931.2022.2054781
40. Ishiguro H (2006) Android science: conscious and subconscious recognition. Connect Sci. https://doi.org/10.1080/09540090600873953
41. Kahn PH, Friedman B, Perez-Granados DR, Freier NG (2004) Robotic pets in the lives of preschool children. Interact Stud 7(3):405–436. https://doi.org/10.1075/is.7.3.13kah

42. Kayukawa Y, Takahashi Y, Tsujimoto T, Terada K, Inoue H (2017) Influence of emotional expression of real humanoid robot to human decision-making. In: IEEE International Conference on Fuzzy Systems (FUZZ-IEEE), 1–6. https://doi.org/10.1109/FUZZ-IEEE.2017.8015598
43. Krach S, Hegel F, Wrede B, Sagerer G, Binkofski F, Kircher T (2008) Can machines think? Interaction and perspective taking with robots investigated via fMRI. PLoS ONE 3(7):1. https://doi.org/10.1371/journal.pone.0002597
44. Kwak SS, Kim Y, Kim E, Shin C, Cho K (2013) What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. IEEE RO-MAN 2013:180–185. https://doi.org/10.1109/ROMAN.2013.6628441
45. Lee KM, Jung Y, Kim J, Kim SR (2006) Are physically embodied social agents better than disembodied social agents? The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction. Int J Hum Comput Stud. https://doi.org/10.1016/j.ijhcs.2006.05.002
46. Lee SA, Liang Y (2016) The role of reciprocity in verbally persuasive robots. Cyberpsychol Behav Soc Netw 19(8):524–527. https://doi.org/10.1089/cyber.2016.0124
47. Li J (2015) The benefit of being physically present: a survey of experimental works comparing copresent robots, telepresent robots and virtual agents. Int J Hum Comput Stud 77:23–37. https://doi.org/10.1016/j.ijhcs.2015.01.001
48. Lim V, Rooksby M, Cross ES (2020) Social robots on a global stage: establishing a role for culture during human–robot interaction. Int J Soc Robot. https://doi.org/10.1007/s12369-020-00710-4
49. Magee JC, Langner CA (2008) How personalized and socialized power motivation facilitate antisocial and prosocial decision-making. J Res Pers 42(6):1547–1559. https://doi.org/10.1016/j.jrp.2008.07.009
50. Moisan F, ten Brincke R, Murphy RO, Gonzalez C (2018) Not all Prisoner's Dilemma games are equal: incentives, social preferences, and cooperation. Decision. https://doi.org/10.1037/dec0000079
51. Munafò MR (2016) Open Science and Research Reproducibility. Ecancermedicalscience 10:1. https://doi.org/10.3332/ecancer.2016.ed56
52. Murphy RO, Ackermann KA (2015) Social preferences, positive expectations, and trust based cooperation. J Math Psychol. https://doi.org/10.1016/j.jmp.2015.06.001
53. Murphy RO, Ackermann KA, Handgraaf M (2011) Measuring Social Value Orientation. SSRN 6(8):771–781. https://doi.org/10.2139/ssrn.1804189
54. Ostrowski AK, DiPaola D, Partridge E, Park HW, Breazeal C (2019) Older adults living with social robots: promoting social connectedness in long-term communities. IEEE Robot Autom Mag 26(2):59–70
55. Pu L, Moyle W, Jones C, Todorovic M (2019) The effectiveness of social robots for older adults: a systematic review and meta-analysis of randomized controlled studies. Gerontologist 59(1):e37–e51
56. R Core Team (2020) R: a language and environment for statistical computing [Internet] (4.0.0). Foundation for Statistical Computing
57. Rapoport A (1967) A note on the "index of cooperation" for Prisoner's Dilemma. J Conflict Resolut. https://doi.org/10.1177/002200276701100108
58. Riek LD, Adams A, Robinson P (2011) Exposure to cinematic depictions of robots and attitudes towards them. The Role of Expectations in HRI
59. Ruijten PAM, Haans A, Ham J, Midden CJH (2019) Perceived human-likeness of social robots: testing the rasch model as a method for measuring anthropomorphism. Int J Soc Robot. https://doi.org/10.1007/s12369-019-00516-z
60. Sandoval EB, Brandstetter J, Obaid M, Bartneck C (2016) Reciprocity in human-robot interaction: a quantitative approach through the Prisoner's dilemma and the ultimatum game. Int J Soc Robot 8(2):303–317. https://doi.org/10.1007/s12369-015-0323-x
61. Schrempf OC, Hanebeck UD, Schmid AJ, Worn H (2005) A novel approach to proactive human-robot cooperation. ROMAN 2005 IEEE International Workshop on Robot and Human Interactive Communication 2005:555–560. https://doi.org/10.1109/ROMAN.2005.1513838
62. Seo SH, Geiskkovitch D, Nakane M, King C, Young JE (2015) Poor Thing! Would You Feel Sorry For a Simulated Robot?: A Comparison of Empathy toward a Physical and a Simulated Robot. Proceedings of the 10th Annual ACM/IEEE International Conference on Human–Robot Interaction - HRI '15. https://doi.org/10.1145/2696454.2696471
63. Straßmann C, Rosenthal-von der Pütten AM, Krämer NC (2018) With or against Each Other? The Influence of a Virtual Agent's (Non)cooperative Behavior on User's Cooperation Behavior in the Prisoners' Dilemma. Adv Hum–Comput Interaction 2018:1–7. https://doi.org/10.1155/2018/2589542
64. Swanson DL (1996) Neoclassical economic theory, executive control, and organizational outcomes. Human Relations 49(6):735–756. https://doi.org/10.1177/001872679604900602
65. Syrdal DS, Dautenhahn K, Koay K, Walters ML (2009) The negative attitudes towards robots scale and reactions to robot behaviour in a live human-robot interaction study. Adapt Emergent Behav Complex Syst. https://doi.org/10.1157/13126291
66. Terada K, Takeuchi C (2017) Emotional expression in simple line drawings of a robot's face leads to higher offers in the ultimatum game. Front Psychol 8:724. https://doi.org/10.3389/fpsyg.2017.00724
67. Van Lange PAM, Joireman J, Van Dijk E (2013) The psychology of social dilemmas: a review. Organ Behav Hum Decis Process 120(2):125–141. https://doi.org/10.1016/J.OBHDP.2012.11.003
68. van Straten CL, Peter J, Kühne R (2020) Child-robot relationship formation: a narrative review of empirical research. Int J Soc Robot 12(2):325–344. https://doi.org/10.1007/s12369-019-00569-0
69. Wilson W, Wong J (1968) Intergroup attitudes towards cooperative vs competitive opponents in a modified prisoner's dilemma game. Perceptual Motor Skills 27(3):1059–1066. https://doi.org/10.2466/pms.1968.27.3f.1059
70. Wu J, Paeng E, Linder K, Valdesolo P, Boerkoel JC (2016) Trust and cooperation in human–robot decision making. The 2016 AAAI Fall Symposium 16(1):110–116. https://doi.org/10.1111/j.1835-2561.2006.tb00045.x
71. Yu O, Aikawa H, Shimomura K, Kondo H, Morishima A, Hun-ok Lim, Takanishi A (2006) Development of a new humanoid robot WABIAN-2. In: Proceedings 2006 IEEE international conference on robotics and automation, 2006. ICRA 2006, pp 76–81. https://doi.org/10.1109/ROBOT.2006.1641164

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Te-Yi Hsieh is currently a post-doc at Chung Shan Medical University (Taiwan). She holds a PhD in Neuroscience and Psychology from the University of Glasgow (UK). Through her doctoral research, she investigated human¬robot cooperation through a psychology and game theory approach in the Social Brain in Action (SoBA) Laboratory. In particular, her research aims to identify personal, robotic, and situational factors that promote cooperative and social relationships between people and physically embodied robots in real life. Additionally, she is interested in understanding how emotional displays of artificial agents influence people's decision-making process.

Bishakha Chaudhury is currently an associate software developer at JPMorgan Chase & Co. She completed a BSc in Computer Science from Fergusson College (Pune, India), a Masters in Computer Application from Chennai University (India) and a Masters in Virtual Reality and Computer Graphics from the University Of Sussex (UK). All the computer games she played while growing up piqued her interest in virtual reality, computer gameplay and artificial intelligence. Her masters dissertation and the game mods and demos she madeled her to a career in the games industry in London. She worked mainly as an AI programmer and was responsible for AI racers and developing variety in the characteristics of non-human players. After this, she joined a team working on computer-aided orthopaedic surgery, where a robotic hand was used for performing precision hip and knee surgery. She then followed her family to scenic North Wales and worked in the local IT industry for six years as a senior developer and technical analyst. She then worked with the Social Robots team as part of the Social Brain in Action (SoBA) Lab for several years across both Bangor and Glasgow Universities.

Emily S. Cross directs the Social Brain in Action Laboratory, which has recently relocated to ETH Zurich in Switzerland. As a Professor of Cognitive and Social Neuroscience, Emily leads a team that uses behavioural and brain imaging tools to explore how different kinds of experience (including lifespan development, culture, and laboratory-induced training) shape how we perceive other agents and actions we encounter in a social world. She completed a PhD in cognitive neuroscience at Dartmouth College following earlier training in dance and psychology. She has previously held faculty positions at Radboud University Nijmegen, Bangor University, the University of Glasgow, Macquarie University and Western Sydney University.