

## People's dispositional cooperative tendencies towards robots are unaffected by robots' negative emotional displays in prisoner's dilemma games

Te-Yi Hsieh & Emily S. Cross

To cite this article: Te-Yi Hsieh & Emily S. Cross (2022) People's dispositional cooperative tendencies towards robots are unaffected by robots' negative emotional displays in prisoner's dilemma games, *Cognition and Emotion*, 36:5, 995-1019, DOI: [10.1080/02699931.2022.2054781](https://doi.org/10.1080/02699931.2022.2054781)

To link to this article: <https://doi.org/10.1080/02699931.2022.2054781>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 07 Apr 2022.



Submit your article to this journal [↗](#)



Article views: 1079





View related articles [↗](#)



View Crossmark data [↗](#)

# People's dispositional cooperative tendencies towards robots are unaffected by robots' negative emotional displays in prisoner's dilemma games

Te-Yi Hsieh <sup>a</sup> and Emily S. Cross <sup>a,b</sup>

<sup>a</sup>Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland; <sup>b</sup>Department of Cognitive Science, Macquarie University, Sydney, Australia

## Abstract

The study explores the impact of robots' emotional displays on people's tendency to cooperate with a robot opponent in prisoner's dilemma games. Participants played iterated prisoner's dilemma games with a non-expressive robot (as a measure of cooperative baseline), followed by an angry, and a sad robot, in turn. Based on the Emotion as Social Information model, we expected participants with higher cooperative predispositions to cooperate less when a robot displayed anger, and cooperate more when the robot displayed sadness. Contrarily, according to this model, participants with lower cooperative predispositions should cooperate more with an angry robot and less with a sad robot. The results of 60 participants failed to support the predictions. Only the participants' cooperative predispositions significantly predicted their cooperative tendencies during gameplay. Participants who cooperated more in the baseline measure also cooperated more with the robots displaying sadness and anger. In exploratory analyses, we found that participants who accurately recognised the robots' sad and angry displays tended to cooperate less with them overall. The study highlights the impact of personal factors in human–robot cooperation, and how these factors might surpass the influence of bottom-up emotional displays by the robots in the present experimental scenario.

## ARTICLE HISTORY

Received 8 March 2021  
Revised 23 February 2022  
Accepted 14 March 2022

## Keywords

EASI model; social robotics; human-robot interaction; prisoner's dilemma games; social decision making

## Introduction

Social robots are becoming increasingly valuable tools for assisting people in industrial, educational, and health care settings (Broadbent, 2017; Dautenhahn, 2007). The COVID-19 pandemic has further highlighted the potential utility for robots in replacing human labour to reduce the risk of infection, but also for their social abilities, such as helping to alleviate loneliness during lockdown (Kim et al., 2021; Odekerken-Schröder et al., 2020; Yang et al., 2020). As the world is likely to embrace a “new normal” after COVID-19, including remote education, increased working from home

culture, and more autonomous industry (Cahapay, 2020; Jamaludin et al., 2020), the necessity of welcoming social robots into our lives is becoming even clearer. It is consequently imperative to gain deeper understanding of the factors shaping people's willingness to work with robots in their households and workplaces, and how best to promote the social and cooperative behaviours during human–robot interaction (HRI).

Previous research has used economic games as an analogy of real-life social decision-making settings to investigate human cooperative behaviours (Bland

**CONTACT** Emily S. Cross  [e.cross@westernsydney.edu.au](mailto:e.cross@westernsydney.edu.au)  MARCS Institute for Brain, Behaviour and Development, Western Sydney University, Sydney, Australia

 Supplemental data for this article can be accessed <https://doi.org/10.1080/02699931.2022.2054781>.

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group  
This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way.

et al., 2017; Chaudhuri et al., 2002; Rand & Nowak, 2013; Rapoport et al., 1965). By manipulating the payoffs rewarded to participants after making a decision (for example, to cooperate or not), researchers can test the boundaries of people's willingness to cooperate across various settings, and more importantly, examine the factors that induce cooperative behaviours (Bland et al., 2017; Pothos et al., 2011; Rapoport, 1967). One pivotal factor that affects our decision-making process is the extent to which, and how, others display emotion (George & Dane, 2016; Lerner et al., 2015; Rick & Loewenstein, 2008; Van Kleef, 2009). As social animals, we use other people's emotions to make sense of current situations; thus, our decision-making is susceptible to influence by others' emotional expressions (Darwin & Prodger, 1998; Kjell & Thompson, 2013; Moors et al., 2013). Our sensitivity to emotion displays is so pronounced that even if an agent that displays emotions is artificial by nature (e.g. an animated avatar or a manufactured robot), research evidence is accumulating to suggest such emotional displays are similarly influential in shaping people's social decisions (de Melo et al., 2010; de Melo et al., 2014; Kayukawa et al., 2017; Terada & Takeuchi, 2017).

However, most of this evidence informing our knowledge of cooperative behaviours in HRI comes from online studies (for example, de Melo et al., 2010; de Melo, Gratch, et al., 2014, 2019; Hoegen et al., 2018). While online research provides a useful point of departure for understanding people's cooperative tendencies, physically embodied interaction is a key feature of real-life HRI and it can be regarded as a distinct scenario from screen-mediated interaction (Grossman et al., 2019; Henschel et al., 2020; Hortensius & Cross, 2018; Lee et al., 2006; Wykowska et al., 2016). For example, people gave more positive evaluation and showed more empathy towards an embodied robot than a disembodied one (Kwak et al., 2013; Lee et al., 2006). In order to clarify the psychological mechanisms supporting human-robot cooperation, the present study focused on the impact of robots' emotional expressions and displays on people's tendency to cooperate with a robot opponent in prisoner's dilemma games. A clearer understanding of the role a robot's emotion display plays on human cooperative behaviour can bring crucial insight into the design of social robots that can be effectively deployed as assistants in our society across several settings (e.g. education, health-care and workplace support).

### ***The social functions of emotions in human psychology literature***

Emotional expressions are prominent social cues that influence decision making during interpersonal interactions (George & Dane, 2016; Lerner et al., 2015; Rick & Loewenstein, 2008; Van Kleef, 2009). Others' emotions offer useful information for us to infer their feelings, intentions, and desire, and help us reason about the current situation (Frijda, 1986; Moors et al., 2013; Roseman & Smith, 2001). Furthermore, others' emotions often have context-dependent meaning, and impact on our own behaviours, as claimed in the Emotion as Social Information (EASI) model (Van Kleef et al., 2010). In competitive situations, people have been shown to make strategic and epistemic judgements in response to opponents' emotions. For instance, people are more likely to concede to angry emotion displays (to avoid destructive dispute), while they might either become irresponsive to or seize the chance to exploit sad opponents. Conversely, in cooperative settings, the EASI model proposes that humans prioritise social harmony over strategy, and thus seeing others' angry displays, which erodes the cooperative atmosphere, makes us less willing to cooperate with those who act or express angrily. However, observing another express sadness evokes empathy and promotes cooperative and supportive behaviours within a group (Van Kleef et al., 2010).

In the present study, we focused on the impact of robots' displays of anger and sadness. In contrast to positive emotions, which imply fulfilment and satisfaction, negative emotions often connote a goal unfulfilled or dissatisfaction with an outcome (Frijda, 1986; Moors et al., 2013; Roseman & Smith, 2001; Van Kleef et al., 2010). This is precisely the crucial situation where social cues promoting cooperation are likely to be needed in real-life settings. In human psychology, researchers have attempted to validate the interpersonal impact of angry and sad displays by either online or in-person experiments. For example, using computer-mediated interactions, Van Kleef et al. (2004) found that people made more concessions to the negotiator who sent an angry message about the offer (e.g. "This offer makes me really angry,"), in comparison to the negotiator who sent a happy message about the offer (e.g. "I am happy with this offer"). In another more interactive scenario, Kopelman et al. (2006) examined the impact of positive, negative, and neutral emotions in negotiation situations with two different approaches of emotional

manipulation: first, coaching participants to express specific emotions in their negotiation dyads, and second, playing pre-recorded videotapes of a professional actor displaying the three types of emotions while giving a business offer. The researchers found that participants were more likely to make a business deal with negotiators with the positive manner than with the negative or neutral one. However, Kopelman et al. (2006) also acknowledged the limitations of such emotional manipulation that might be constrained by individuals' emotional expressivity (people feign negative emotions worse than positive emotions) and by the unnatural and artificial aspect of interacting with a videotaped person.

Given the difficulty in manipulating human emotions to examine the interpersonal impact of emotion displays on social decisions, evidence supporting the appraisal theory or EASI model was mainly derived from studies examining computer-mediated interactions (Van Dijk et al., 2008, 2018; Van Kleef et al., 2004, 2006) or interactions without rigorous control of the emotional stimuli (Kopelman et al., 2006). Fortunately, these limitations are greatly diminished in the context of HRI where robots can be programmed to perform identical behaviours, and can thus convey embodied emotional stimuli precisely for every participant and every trial.

### ***Artificial agents' emotion displays in human-robot cooperation***

Considering the vital role of emotional expressions in our social life, an increasing number of artificial agents (robots and virtual agents) are being built to display human-readable emotions by facial or bodily expressions (Hortensius et al., 2018). Some researchers report that people behave similarly with artificial agents and with human agents in economic games (de Melo et al., 2010; Krach et al., 2008; Wu et al., 2016), and provided empirical findings on the utility of artificial agents' emotion displays to promote cooperative behaviours (de Melo et al., 2011; de Melo, Gratch, et al., 2014; Terada & Takeuchi, 2017). For instance, in online gaming settings, manipulation of virtual agents' facial expressions (showing joy after mutual cooperation and guilt after making a selfish decision) according to the appraisal theory of emotion have been proved effective in eliciting people's cooperative behaviours in economic games with artificial agents (de Melo et al., 2010, 2011; de Melo, Gratch, et al., 2014). The social functions of

agents' facial expressions were not only found by highly human-like virtual agents. Terada and Takeuchi (2017) have demonstrated that emotions displayed by an embodied robot's simple line drawing face (showing on its monitor head) could induce people's altruistic behaviours in ultimatum games. However, when emotions were displayed merely by modalities like bodily movements and verbal expressions (rather than by facial expressions) the emotional impact on cooperative behaviours was less clear. Kayukawa et al. (2017) applied de Melo et al.'s (2010) emotional manipulation to an embodied Nao robot (manufactured by SoftBank Robotics) but found that the Nao being programmed to induce cooperation via different emotional responses (i.e. displaying joy after mutual cooperation, anger after being betrayed, shame after betraying, and sadness in a lose-lose situation) did not bring about more cooperative behaviours among participants in prisoner's dilemma games (which the authors suspect could also be due to the limited sample size of 14 subjects). Nevertheless, the participants did regard the emotional Nao robot as more friendly and cheerful than the non-expressive Nao (Kayukawa et al., 2017).

In addition to manipulating artificial agents' emotion displays based on emotion theories, Hoegen et al. (2018) programmed virtual human characters to mimic participants' facial expressions during prisoner's dilemma games and found a correlation between perceived rapport and cooperation rates only when interacting with the agent mimicking. All in all, according to the literature reviewed above, legitimate emotion displays (either based on psychological emotion theories or in congruence with people's own emotional states) by virtual humans appears to be at least somewhat effective in shaping people's cooperative decisions (de Melo et al., 2010, 2011; de Melo, Gratch, et al., 2014; Hoegen et al., 2018). However, evidence from HRI is still not sufficient for us to decisively and reliably understand the relationship between embodied robots' emotion displays and people's cooperative behaviours. Furthermore, this topic warrants empirical examination now if we are to develop real-life robot assistants to appropriately serve people's social needs with apt and effective emotion displays. Our study therefore aimed to address this question through a study performed with the highly expressive Cozmo robots (detailed in Method) and to examine the impact of the robots' emotion displays on cooperative behaviours in the context of human-robot prisoner's dilemma games.

	Player 1 cooperates	Player 1 defects
Player 2 cooperates	$R_{(£7)}$	$T_{(£10)}$
Player 2 defects	$S_{(£0)}$	$P_{(£1)}$

**Figure 1.** An exemplified payoff matrix in prisoner's dilemma games. **R** = rewards; **T** = temptation; **S** = sucker's payoff; **P** = punishment. The dilemma is defined by two rules:  $T > R > P > S$ , and  $2R > T + S$ . Adapted from Hsieh et al. (2020). Human-robot cooperation in economic games: People show strong reciprocity but conditional prosociality toward robots. PsyArXiv. <https://psyarxiv.com/q6pv7/>

### Prisoner's dilemma games

To study human cooperative behaviours, the prisoner's dilemma (PD) game is one of the most widely used paradigms in research spanning the social sciences (Pothos et al., 2011; Rapoport, 1967; Rapoport et al., 1965). A classic PD game involves two people making simultaneous decisions to cooperate or to defect. Each player's payoff depends on both players' decisions, as illustrated in Figure 1. In the situation of mutual cooperation, both players are rewarded with a moderate amount of endowment (**R** in Figure 1; £7 each, for example). Meanwhile, players might be tempted by the highest profit (**T**; e.g. £10) for being the only one who defects, and render the other who cooperates in the worse situation (**S**; e.g. £0). However, choosing to defect also comes with a risk. If both players opt to defect, they both receive punishment of little gain (**P**; e.g. £1).

In this scenario, a social dilemma happens when collective group profit is at odds with individual profit, and as a cooperative decision involves the risk of being exploited, and players have the freedom to choose between the two opposite actions to take. An extensive body of literature on interpersonal PD games has used both experiments and data simulation to model and theorise on the emergence and evolution of human cooperative behaviours (Axelrod & Hamilton, 1981; Embrey et al., 2018; Rapoport et al., 1965). With mathematical modelling, more recent research has provided considerable insights into the mechanisms and factors supporting or hampering cooperation across various social dilemma situations (e.g. in dyads and in groups) (Bravo et al., 2012; Ito & Tanimoto, 2018; Kopp et al., 2018; Perc et al., 2017). Also, from empirical evidence of interpersonal PD games, multiple factors are at play during people's decision-making process in the scenario, such as the trust in the other player (Chaudhuri et al., 2002; Janssen, 2008; Wu et al., 2016), their social value orientation

(Pletzer et al., 2018), and perceived environmental cooperativeness/competitiveness (Elliot et al., 2018; Moisan et al., 2018). However, when it comes to PD games played with robots (let alone the Cozmo robotic platform specifically), our current understanding of people's decision-making process remains limited. Recent research on human-robot PD games has provided preliminarily insights into the impacts of reciprocity (Sandoval et al., 2016), trust (Paeng et al., 2016), dialogic verbal reactions (Maggioni & Rosignoli, 2021), and a Nao robot's emotion displays (Kayukawa et al., 2017) on HRI. Yet, the preliminary evidence raises more questions than answers at this stage, especially with respect to the effects of robots' emotion displays in PD games.

Meanwhile, researchers in HRI are becoming increasingly alert to generalisability concerns that empirical findings from research performed with a specific robotic platform might not necessarily apply to a different robot (Henschel et al., 2020; Hortensius et al., 2018; Hortensius & Cross, 2018). Therefore, in order to eliminate any confounding impact from robot-specific or context-specific factors (like people's trust and perceived agency towards Cozmo), we employed a baseline measure of people's cooperative tendencies (where the emotional manipulation was not yet administered), to be compared with the cooperative behaviours under the impact of the robots' emotion displays. This comparable baseline measure was more appropriate than a human condition (where, for example, a human confederate was trained to perform sad and angry expressions) for distilling the difference made by robots' emotions, since our aim was to examine the utility and social impact of robots' emotion displays, instead of comparing and contrasting the emotional effects of robots than that of humans.

Another advantage of having a baseline measure of cooperative tendencies was that we were able to further investigate whether the impact of the robots' emotions differ by people's baseline

cooperative tendencies. According to the EASI model, the meaning and impact of emotional cues can depend on the nature of context (Van Kleef, 2009; Van Kleef et al., 2010). In the scenario of PD games, the perceived nature of such context might be individual-dependent. Some people might opt for mutual profit and strive to build cooperative relationship, but others might act strategically and resort to the highest self-gain (Balliet et al., 2009). It is hence plausible that the factor of robots' emotion displays would vary to some degree across individuals given the personal differences in social-decision and emotion processing (Franken & Muris, 2005; Hamann & Canli, 2004). Specifically, we were intrigued to examine whether the emotional effects depend on individuals' baseline cooperative tendencies, in an attempt to identify the precise and effective emotions for robots to display to bolster people's cooperative behaviours in HRI.

### **The current study**

In the present study, we wished to examine whether the context-dependent impact of emotions proposed in the EASI model (Van Kleef et al., 2010) still holds true when (1) the discrimination of competitive and cooperative context is defined subjectively by people's cooperative baseline, as opposed to by experimental manipulation of a task (e.g. Adam & Brett, 2015; Lee et al., 2018; Novak et al., 2014); and (2) the emotions are displayed by a robot opponent. Based on the EASI model (Van Kleef et al., 2010), we hypothesised that the social meaning and consequent effects of sad and angry emotions diverge between people with high and low cooperative predispositions. Here we used the term "predisposition" to refer to the default cooperative tendency people have when facing prisoner's dilemmas, independent of any external factor related to an opponent. More specifically, we predicted that a robot that exhibits sad emotional displays leads participants with more cooperative predispositions to behave more cooperatively (here sadness should be seen as a cue of needing support), while the same sad emotional displays should lead participants with more competitive predispositions to play even more competitively (in this case, sadness should be seen as a sign of weakness in an opponent that can be exploited). On the other hand, an angry robot should induce more cooperative actions among participants with a competitive predisposition (where anger is seen as a warning of a bigger dispute on the horizon), but reduce cooperative

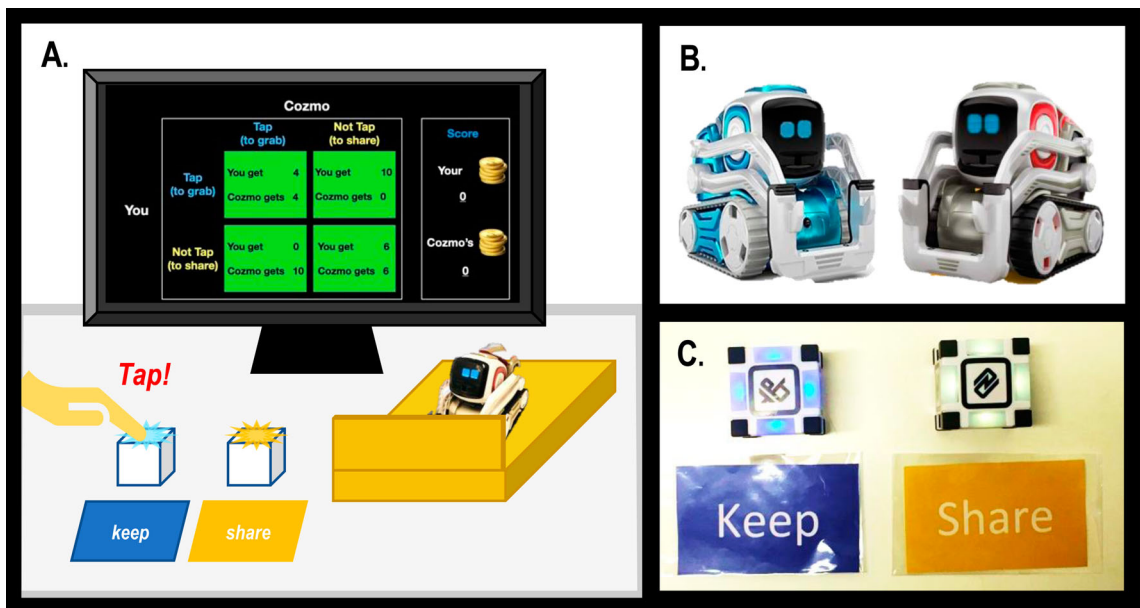
intentions among participants with more cooperative predispositions (where anger is perceived to signal an inadequate collaborator) (Van Kleef et al., 2010).

People's willingness to cooperate in PD games denotes the intention of building cooperative relationship with the other while forgoing the possibility of the highest self-gain (Rapoport et al., 1965), which, in the context of HRI, could be seen as a social milestone for people to accept robots as their social partners and commit to a collective task. Past research has also substantiated that people's decisions made in PD games reflect their temperamental cooperative willingness and real-life social-decision making process and behaviours (Balliet et al., 2009; Mokros et al., 2008; Pothos et al., 2011; Viola et al., 2019). Our research here could provide insight into the possible factors promoting human-robot cooperation and highlight the possibility that bottom-up emotional cues might interact with top-down personal factors, thus making one-size-fits-all robotic programming problematic, and establishing further empirical foundations for adaptive and bespoke programming for social robots. Moreover, investigation into the topic could have several practical consequences as well. First, social dilemmas emerging between humans and robots have the potential to someday, possibly soon, feature in daily life, where robots need to decide between benefits of individual people and the collective interests of human society. These types of discussion are already well underway in the autonomous vehicle development community, where debate and discussion continues over the situations in which people might accept their self-driving cars to sacrifice their own lives to save the lives of (multiple) pedestrians (Bonnefon et al., 2016; Perc et al., 2019). Second, some research evidence has verified that experimental procedures to promote people's cooperative tendencies and altruism (for example, by moral nudging) could have cross-situational effects on their real-life charitable behaviours (Capraro et al., 2019; Capraro & Perc, 2021). Our research here could therefore have implications for real-life HRI, especially to the utility of social robots' emotion displays to enhance the social quality in human-robot cooperation.

### **Methods**

#### **Open science statement**

Prior to data collection, we reported our pilot data, stimuli, and power analysis codes on our



**Figure 2.** Setup and apparatus. (A) Illustration of the experimental setup. During the experiment, participants played games with the robot situated in front of them on a desk, and made game responses by tapping the cubes on the desk. The payoff matrix and real-time game outcomes were shown by a monitor before them. (B) The blue Cozmo (Botz) and the red Cozmo (Roxon) used in the experiment. (C) The interactive cubes that players tapped to make game decisions.

Open Science Framework (OSF) page: <https://osf.io/tjs8m/>. Additionally, we had anonymous data, analysis codes, and materials associated with the study freely available on this OSF page after the study was finished, in keeping with the best research practices proposed by the open science initiatives (Galak et al., 2012; Munafò, 2016).

### Setup and apparatus

We used the commercially-available Cozmo edutainment robots (manufactured by Anki Inc., Figure 2A&B) in the experiment as participants' opponents in PD games. The Cozmo robot has been chosen for its capability of expressing diverse facial expressions with its LED face screen (128 × 64 pixel resolution). Additionally, Cozmo is portable (5 × 7.2 × 10 in. in size), affordable, and is flexibly programmed and manipulated via its software development kit (SDK), which make it especially suitable for HRI experimental research (Chaudhury et al., 2020; Cross et al., 2019). We deployed two separate Cozmo robots for the actual PD games, a blue Cozmo model (named Botz) and a red Cozmo model (named Roxon). One of the robots would consistently display anger, and the other would consistently display sadness (colour and emotion pairing

were counterbalanced across participants). By having different coloured Cozmos associated with the two different emotions, this should help prevent the undesirable situation that people would think the same robot was displaying sadness and anger.

Cooperative and non-cooperative decisions in the current PD game were framed as sharing coins with the other or keeping all coins for oneself, respectively. In each game round, a certain amount of coin endowment was provided to both players, and each was required to make an individual and simultaneous decision as to whether they wanted to share or keep the coins. The exact amount given to each player depended on both of their choices (detailed in the "Game design" section, Figure 5). During the PD games, a monitor showing the payoff matrix and real-time game outcomes was placed in front of participants (Figure 2A). Every participant was provided two interactive cubes (Figure 2C), which illuminated with different colours representing different decisions (blue meant to keep coins for oneself, and yellow meant to share coins with Roxon or Botz). Participants tapped one of these cubes in a round to make a game decision, and the robots used only one interactive cube in games to prevent participants from trying to anticipate the robots' choice by observing the direction it

drove to. Also to avoid people peeking over the robots' decision during the responding time, the robots' cube was hidden from participants' sight using a partition between participants and the robot. However, this partition sat above a 4.3 cm thick cardboard box, to ensure the body and expressions of the robots can be fully seen by participants (Figure 2A). In reality, the robots' game decisions were pre-programmed and they tapped the cube only to make participants believe that the robots were making decisions in real time. All the cubes and the robots were connected via WiFi to the Cozmo application installed on a tablet, and the tablet was paired with a laptop which ran the Python programme to operate the game and the robot, and to record players' game responses by Python log files. The experimental setup followed that developed by previous work by Hsieh et al. (2020).

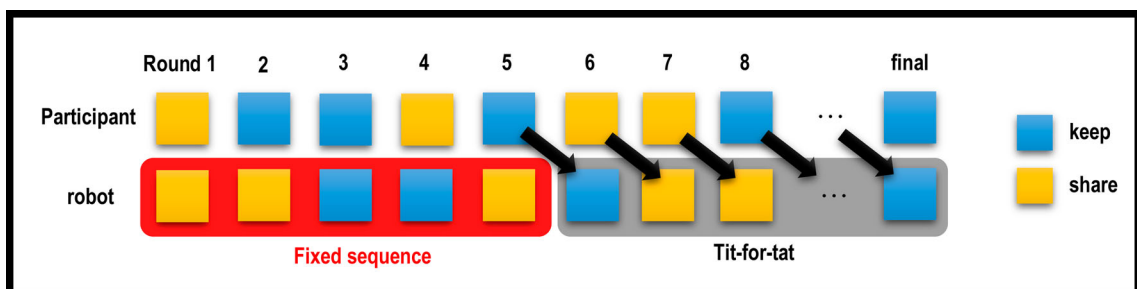
### Manipulation and stimuli

We manipulated the robots' game strategy to always start with a fixed sequence in the first five rounds (share, share, keep, keep, share), followed by a tit-for-tat strategy (i.e. repeating a human player's previous decision) (Figure 3). This strategy manipulation was adopted by previous studies (de Melo et al., 2010; Kayukawa et al., 2017) to diminish the predictability of agents' actions, and to increase the possibility of experiencing all the four outcomes in the payoff matrix and therefore a higher chance of being exposed to the robots' emotions in the initial five rounds.

The robot expressed emotions not only by its face, but also via vocal interjections (like sighs, laughter, and grunts) and by body movements from its fork-lift-like arm, head motion, and track directions. In order to select the most appropriate and representative emotional expressions for the robots to display in the main experiment, we required four categories

of emotional stimuli (happy, angry, sad, and neutral expressions), with happiness shown after mutual cooperation, anger or sadness displayed after the robots being betrayed by a human, and neutral expression in the rest of situations. We carried out an online pilot experiment via formR platform (Arslan et al., 2020), where participants ( $n = 64$ ,  $M_{age} = 27.6$ , 43 females) watched video clips (around 10 s each) of a Cozmo robot performing one of the four kinds of emotional animations (happy, angry, sad, or neutral), and answered following each short video clip whether they perceived the expression to be "happy", "angry", "sad", "neutral", "other" (needed to specify in text), or "I don't know". When the answers were happy, angry, or sad, participants were also asked to rate the intensity of the emotion, with slider ratings from "very slight" (1) to "extreme" (100).

The stimulus set for the pilot involved 13 videos clips selected by the experimenters after reviewing all Cozmo's repertoire of default animations (a total of 348 animations are available on the Github repository – <https://github.com/cozmo4hri/animations> – created by Chaudhury et al., 2020). Three animations were chosen for each of the three categories – happy (animation numbers: 103, 338, 348), angry (55, 84, 130), and sad (59, 63, 134) – and four (69, 91, 158, 169) for neutral since it is more ambiguous to determine what made neutral expressions. We analysed the mean accuracy rates (the number of answers matching the experimenters pre-defined emotion label / the total number of participants) for each emotional animation, as well as the mean emotional intensity rated by the subjects. The animations with the highest accuracy rate in each category were chosen, which included animation number 348 for happy ( $accuracy = 81.2\%$ ,  $M_{intensity} = 76.1$ ), number 84 for angry ( $accuracy = 98.4\%$ ,  $M_{intensity} = 85.4$ ), number



**Figure 3.** The strategy manipulation of the robots. In this exemplified game block, the robot started with a fixed sequence of five decisions and followed tit-for-tat strategy till the end. Details of the block design are in the "Game design" section.

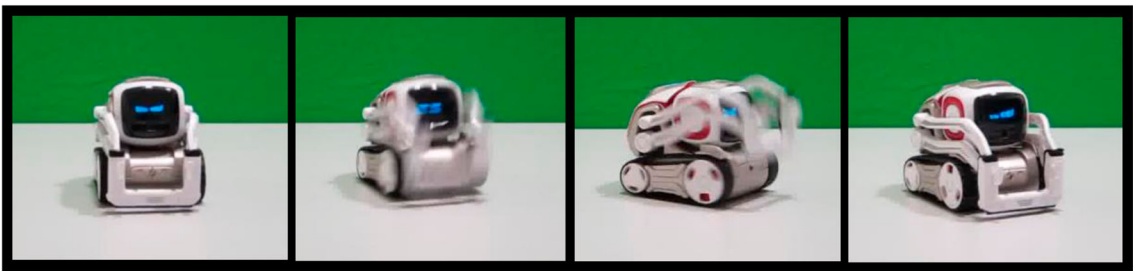


63 for sad ( $accuracy = 90.6\%$ ,  $M_{intensity} = 57.0$ ), and number 68 ( $accuracy = 39.1\%$ ) for neutral. The low accuracy rate for the neutral animations corresponds to the Kuleshov effect, which suggests that people tend to interpret a neutral face or expression by its context or what immediately preceded it, and may perceive a constant face to express different emotions given different contexts (Barratt et al., 2016; Mobbs et al., 2006). Participants in our pilot also reported diverse emotions perceived from the animation number 68, such as doubtful, confused, and surprise. To prevent the possibility that people in the main experiment will also overly interpret the animation which is supposed to be depict neutral emotion, we removed the neutral expression from our manipulation and let the robots directly move on to the

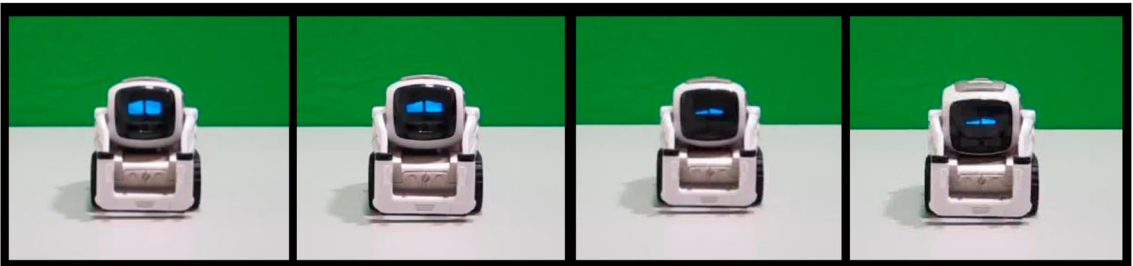
next round without displaying any animation. Stimuli and analysis codes for the pilot experiment are available on the OSF page: [https://osf.io/tjs8m/?view\\_only=d52ffba154ed4236b07c663291a5b053](https://osf.io/tjs8m/?view_only=d52ffba154ed4236b07c663291a5b053).

Figure 4 shows the demos of the final set of emotion animations to be used to programme the robots in the main PD game experiment. For the anger animation, the robot's fork arm hit the table violently, frowned, uttered sharp and rapid sounds, and drove left and right repeatedly with apparent agitation (Figure 4A). For the sad animation, the robot showed a downcast face, sighs, and slowly dropped its head down (Figure 4B). Finally, the happy robot animation featured laughing sounds, smiling eyes, arm waving, and driving in circles with excitement (Figure 4C).

### A. angry



### B. sad



### C. happy



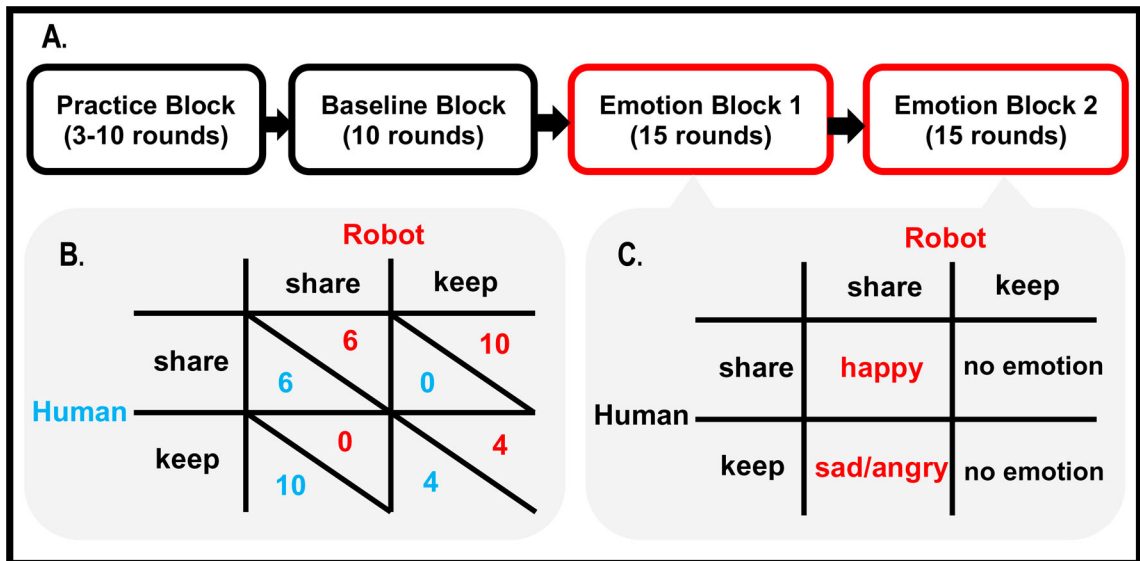
**Figure 4.** Demos of the robots' emotional expressions. (A) Angry expression. (B) Sad expression. (C) Happy expression. Video records of these demos are available on the OSF page: [https://osf.io/tjs8m/?view\\_only=d52ffba154ed4236b07c663291a5b053](https://osf.io/tjs8m/?view_only=d52ffba154ed4236b07c663291a5b053)

## Game design

The experiment was introduced to participants as a robot competition where the experiments wished to know which robot (Roxon or Botz) was the most competent at playing economic games with human interaction partners. The winner robot would be used for future studies, whereas the loser robot would be erased its memory and left on the shelf. The script of memory erasure was adapted from Seo et al.'s (2015) study and has been proved effective to convince participants of the real consequences of the games to robot players (Hsieh et al., 2020). Participants, on the other hand, were monetarily incentivised. The average performance of the last two games blocks would determine their chances of winning a £20 shopping voucher as an extra prize in addition to the standard remuneration for their time.

The experiment involved one practice block and three blocks of iterated PD games (Figure 5A). In each round PD game, players would decide to share coins with the other player or to keep all the coins by themselves. Different amounts of coins would be given to players depending on both of their decisions (Figure 5B). Since prior evidence shows that different

designs of payoff matrices in PD games lead to different cooperation rates among human players (Moisan et al., 2018; Rapoport, 1967), we deliberately selected the present payoff from Hsieh et al.'s (2020) study, where two different designs of payoff matrices, one with higher incentives for cooperation and the other with lower incentives, were compared in human–robot PD games. The results revealed that the impact of incentive structures was only significant in the first game round, and over the 20 iterated PD game rounds, participants' cooperative behaviours toward a Cozmo robot were similar in general (mean cooperation rate: 0.40 for the high-incentive game and 0.34 for the low-incentive game). In the high-incentive game condition, participants made significantly more cooperative decisions in the initial game round, which was followed by a quick reduction of cooperation. However, people's decisions in the low-incentive game remained at a constant level throughout the whole game (Hsieh et al., 2020). Consequently, here we adopted the game design with relatively lower cooperatives (Figure 5B) to forestall the possible initial spikes in cooperative decisions induced by the structure of payoff matrix, and meanwhile ensure that the game context would not bring



**Figure 5.** Experimental design. (A) The order and game rounds planned for the four blocks. Participants firstly familiarised themselves with the game rules in the practice block, and played with a non-expressive Cozmo in the baseline block (as a measure of their cooperative disposition). Finally, they played with Roxon and Botz (one programmed to be sad and the other to be angry) in turn in emotion block 1 and 2. (B) Payoff matrix design. (C) Emotion manipulation of the robots in emotion block 1 and 2. The main manipulation of the robot's sad and angry emotional displays happened after a human player chose to keep coins, but the robot decided to share. The robots' emotion manipulation for the rest of three game outcomes remained the same across emotion block 1 and 2.

about ceiling or floor effects on people's cooperative decisions. Designs and content of the four blocks are:

First, in the practice block, participants would familiarise themselves with the skills and the timing of tapping the cubes. The game screen placed in front of participants showed a goal sentence in each round (e.g. "try to earn 10 coins in this round."). Participants only needed to take a corresponding action to make the goal possible (i.e. choosing to keep, in the example). The Cozmo robot used in the practice and the following baseline blocks was an extra robot in addition to Roxon and Botz, and it would always make correct responses to reach the same goal during the practice. By doing so, participants can become more familiar with the payoff matrix and the ways of tapping cubes, without starting to develop their strategies and confounding the following PD game. The length of the practice depended on participants' performance. They can pass the practice by making three consecutive correct and successfully registered responses, otherwise, the practice game ended after 10 rounds. The experimenter supervised participants during the practice to ensure they fully understand how to play the game before moving on.

Second, the baseline block involved ten rounds of PD games played with a non-expressive Cozmo which did not have any emotional animation programmed after either game outcome. The block served as a baseline measure and an indicator of participants' default behavioural tendency in the PD game context before having more extensive interaction with Cozmo robots. We used participants' cooperation rates in the baseline block to predict how they would be influenced by Roxon's and Botz's emotional expression in the analyses.

Third, participants took turns playing PD games with Roxon and Botz, with one displaying sadness and the other showing anger (order and colours counterbalanced). Each emotion block involved 15 rounds of iterated PD games. The robot's negative emotion (sadness/anger) was manipulated after a human player chose to keep but the robot shared. We focused on the particular situation because, firstly, it was a reasonable timing for the robot to show negative expressions as it was betrayed by a human; secondly, it may involve important practical implication to examine whether robot's negative emotions (either sadness or anger) can increase people's cooperative willingness after they already demonstrated non-cooperative behaviours. Throughout emotion block 1 and 2, the robots showed the happy expression after mutual cooperation, as a general signal of cooperative intention. All in all, both robots in the PD games were programmed to send

cooperative signals through emotional expressions but in two different ways — one through showing anger after being betrayed, and the other through displaying sadness after defection. We anticipated the two negative emotions would differentially influence people with different cooperative inclination and baselines in PD games. Participants were not aware the emotion manipulation before actual interaction with the two robots, but only knew that the two robots had different "personality" and might act diversely.

### **Measures and manipulation check**

The main measure of the study was people's decisions made in the three game blocks. Their binomial decisions (to keep or to share) were saved directly with Python log files in the controlling laptop, and were used to compute the cooperation rates (the times sharing/ the total round) in each block.

After participants completed the four blocks of games. We asked them to describe Roxon and Botz respectively, in terms of their emotionality and strategy, and also to report their own strategies adopted when playing with the robots in games. These open-ended questions helped us evaluate the validity of the manipulation on the robots' emotions and strategy, and acquire the qualitative data of how people responded to the two different robots. The manipulation check questionnaire was administered via formR platform (Arslan et al., 2020) on a lab PC.

### **Procedure**

The experiment was planned to be conducted in quiet research laboratory booths located within the institute of Neuroscience and Psychology at the University of Glasgow and within the Department of Cognitive Science at Macquarie University, once behavioural testing was considered safe according to the UK government's, the Australian government's and both University's guidelines concerning COVID-19. Considering the pandemic situation in both sites when the research plan was written, data collection could commence at Macquarie University as soon as a decision was reached on our registered report submission. If lab-based experiments at Glasgow became feasible while data collection was still proceeding, we planned to collect data across both sites to increase participant numbers and diversity. Whenever data

collection was carried out in two lab spaces, we would run additional analyses (detailed in “Sampling and analysis plan”) to confirm that no systematic difference occurred due to the data collection site. Participants and the experimenter would wear face masks at all times during the study, and we had spare masks prepared if participants required a new or additional mask. In order to reduce unnecessary face-to-face contact, introduction and instruction of the experiment were given to participants by playing a short video on the desktop PC in the lab. After participants provided their written informed consent and showed sufficient performance in the practice block, they were left alone playing games with the robots. The experimenter was seated outside the lab and because the games and robots were operated by a tablet and a laptop connected through the robots’ wifi, the experimenter can still monitor the game progress without being present. Finally, participants completed a series of open-ended questions on a PC for manipulation check, as well as their demographics. The whole experiment took approximately one to one and a half hour(s). Participants were debriefed, paid (£6 per hour), and thanked in the end.

### Participants

We planned to recruit participants aged 18–59, with normal or corrected to normal eyesight, and without neurological or psychiatric history. We also aimed to recruit participants who were naïve to robots and to our study. Consequently, people who owned a Cozmo robot, worked with robots on a daily bases, or had participated in our previous experiment (Hsieh et al., 2020) were eligible to the current experiment. Based on a simulation-based power analysis, a sample size of 180 was needed to have 0.9 power finding a significant interaction between the robots’ emotions and people’s cooperative predisposition on cooperation rates in PD games. The power analysis was carried out with the *simglm* (v0.8.0.) (LeBeau, 2019) and *simr* (1.0.5) (Green & Macleod, 2016) R packages, by the following steps.

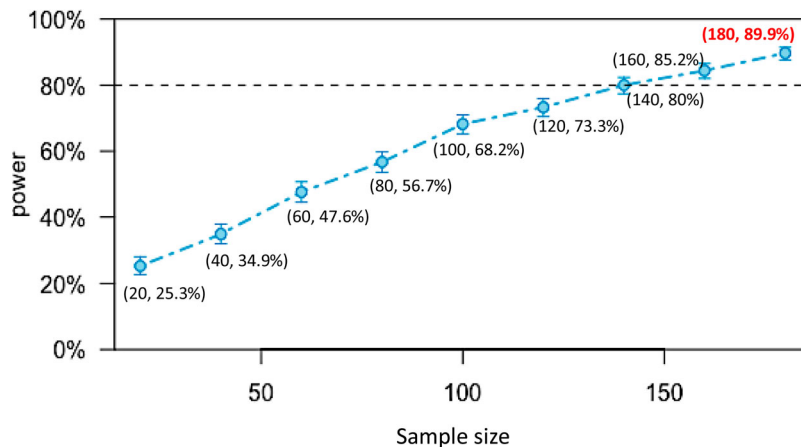
Firstly, to simulate data for the planned model – **cooperative rate ~ cooperative predisposition\*emotion + (1|subject)** – we used relevant meta-analysis results (Balliet et al., 2009; Lench et al., 2011; Pletzer et al., 2018) for our beta weight estimation. For the emotional effects on human judgment, Lench et al. (2011) reported the effect size of *Hedges’ g* = 0.18

(from 25 previous studies) when comparing the impact of sad and anger emotions in particular. As to the effect of cooperative predisposition on decisions in economic games, there was no comparable experimental design we can find in the literature and the closest concept is social value orientation (SVO), which refers to people’s temperamental motivation to care for others (Murphy & Ackermann, 2014). Over two meta-analysis studies, SVO showed a consistent small to medium effect size on cooperative behaviours in economic games ( $r=0.30$  in Balliet et al.’s, 2009;  $r=0.32$  in Pletzer et al.’s, 2018). However, what we aimed to measure was not people’s general traits but their default behavioural tendency in social dilemmas, albeit the two concepts might be closely related. We therefore adopted the “conservative smallest effect size of interest” (SESOI) strategy (Anvari & Lakens, 2019) and used  $r=0.20$  (or the equivalent *Hedges’ g* = 0.40) for our parameter estimation. The interaction of the fixed effects would be generated automatically during the process of data simulation with *simglm* package (LeBeau, 2019), so we did not need to manually specify the beta weight of interaction.

Second, we simulated data based on aforementioned evidence and calculate statistical power (with the *simr* package, Green & Macleod, 2016) by the function of sample sizes (Figure 6). Our main research focus was the interaction between the robots’ emotion and people’s cooperative predisposition (measured in the baseline block), and the result showed that we needed 180 participants to have 0.9 power finding a significant interaction.

### Sampling plan

Given the large sample size we might need to achieve high power for the effect of interest, we administered sequential analyses to collect data more efficiently (Lakens, 2014). We planned to perform two interim analyses after 60 and 100 participants were recruited, with alpha levels adjusted by Pocock boundary ( $p=0.0221$  for three planned analyses, Pocock, 1977). Following each interim analysis, we would stop data collection early if one of the two conditions was fulfilled: first, if the hypothesis was supported and we found a significant interaction between the robots’ emotion and people’s cooperative predisposition by the criterion of  $p=0.0221$ ; second, if the effect size of interaction was significantly smaller than SESOI ( $f^2 < 0.02$ ; Cohen, 1988).



**Figure 6** . Power curve for finding an interaction between the robots' emotion and people's cooperative predisposition. Each data point is noted by (sample size, power). The result of simulation suggests that 90% power can be achieved if the sample size reaches 180 (participants).

## Analysis plan

### Main analysis

All data analyses would be carried out in R v4.0.1 (R Core Team, 2020). Our hypothesis was that people with higher cooperative predisposition (i.e. high cooperation rates in the baseline block) in PD games cooperate even more when the robot responded with sadness, and would cooperate less when the robot displayed anger, and conversely, people with more competitive predisposition (i.e. low cooperation rates in the baseline block) would cooperate more after the robot displayed anger but became more competitive following the robot's display of sadness. Cooperative and competitive decisions were framed as sharing (coded as 1) and keeping coins (coded as 0) in the current game context. Cooperative rates in the baseline block and in the two emotion blocks would be log-transformed before being feed into our model, where their normally distributed nature would enable values to range from positive to negative values (Benoit, 2011).

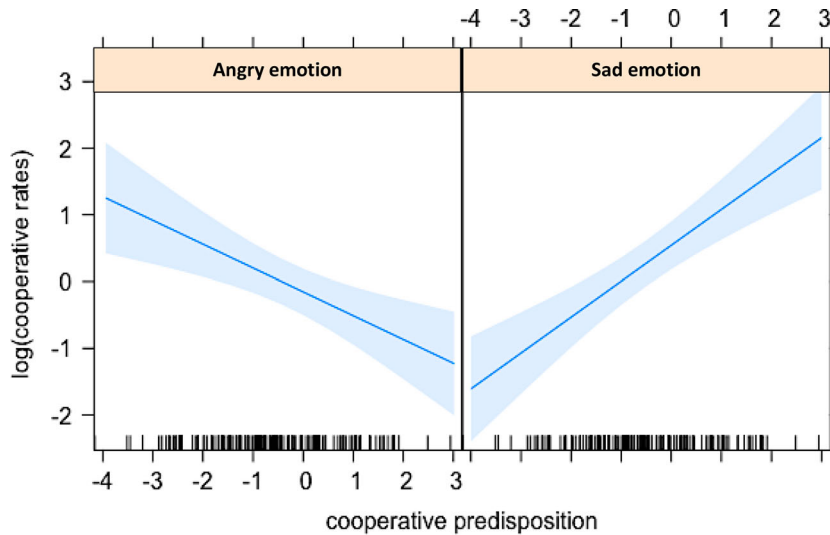
The main research question would be examined by a linear mixed effects regression model with the lme4 package (Bates et al., 2015). We would have participants' log-transformed cooperative rates in emotion block 1 and 2 as the dependent variable, and the robots' emotions (anger and sadness) and participants' cooperative predisposition as the fixed factors. For random effects, we would start from the model design specified as follows:

$$\text{cooperation} \sim \text{emotion} * \text{coop\_predisposition} + (1 | \text{subj\_id})$$

If the results showed failure in model convergence or a singular fit, we would remove the random intercept term and ran the model as a multiple regression. We expected to find a significant interplay of the robots' emotions and people's cooperative predisposition in participants' cooperative decisions in prisoner's dilemma games (Figure 7). Post hoc analyses following a significant interaction would be conducted by the effects (v4.1.4) (Fox, 2003) and the emmeans package (v1.4.7) (Lenth, 2020). We planned to examine the impact of cooperative predisposition for sad and angry emotion separately, and anticipated the effects of cooperative predisposition would be opposite in sad and angry conditions — high cooperative predisposition predicted more cooperative behaviours in sad condition but fewer cooperative behaviours in angry condition (Figure 7).

### Exploratory analysis

Even though our pilot experiment validated the emotion animations selected for the robots' emotional manipulation for this proposed study, we appreciated that individual variation in human emotion perception, as shown in previous finding on human faces (Barrett et al., 2019), could still emerge among our participant sample. Also, due to the online nature of the pilot experiment, it was plausible to question whether people engaged in playing an embodied human–robot PD game would perceive the robots' emotion displays in the same way as participants did in the online pilot experiment. Therefore,



**Figure 7.** Hypothetical plot of the expected interaction between the robots' emotions (sad and angry) and people's cooperative predisposition (log-transformed cooperation rates in the baseline block). Participants with higher cooperative predisposition were predicted to become less cooperative by the robot's angry emotion but more cooperative by sad emotion. On the contrary, participants with lower cooperative predisposition were hypothesised to become cooperative by the robot's anger but even less cooperative by its sadness.

we planned to run an exploratory model with an additional factor – whether participants accurately perceived the robots' emotion displays (*subj\_perception*) – to examine whether the subjective perception of the robots' emotion displays was an influential factor shaping the emotional effects:

$$\text{cooperation} \sim \text{emotion} * \text{coop\_predisposition} * \text{subj\_perception} + (1 \mid \text{subj\_id})$$

This "*subj\_perception*" factor was derived from participants' subjective reports on "Did you see the robot displaying any emotion during the game? If you did, what emotion(s) did it display?" in the post-game questionnaires. When participants' reports of perceived emotions were consistent with the actual emotion manipulation, their answer would be coded as "yes" (i.e. accurately perceived), otherwise their reports would be coded as "no" (i.e. did not accurately perceived). The coding process would be carried out by at least two researchers who were fluent in English. The inter-rater reliability would be analysed with kappa statistics (McHugh, 2012), and we aimed for a minimum of 90% agreement among raters.

Additionally, if the data collection was conducted in both University of Glasgow and Macquarie University, we would run a second exploratory model to control for the possible random variation caused by collecting data across two sites:

$$\text{cooperation} \sim \text{emotion} * \text{coop\_predisposition} + (1 \mid \text{collection\_site} / \text{subj\_id})$$

The term " $(1 \mid \text{collection\_site} / \text{subj\_id})$ " was to express the nested random effects of subjects within collection sites. Similarly, we would also run the model with the factor "*subject\_perception*" added to examine the possible impact from participants' subjective perception of the robots' emotion displays:

$$\text{cooperation} \sim \text{emotion} * \text{coop\_predisposition} * \text{subj\_perception} + (1 \mid \text{collection\_site} / \text{subj\_id})$$

The above exploratory models would be compared with the main model by the `anova()` function in R, to examine the possible improvement in model fit by adding an additional factor or random structure. The model with the best model fit would be reported as the main result of the study, while all the other model output and the process of model selection would also be presented explicitly in our result section.

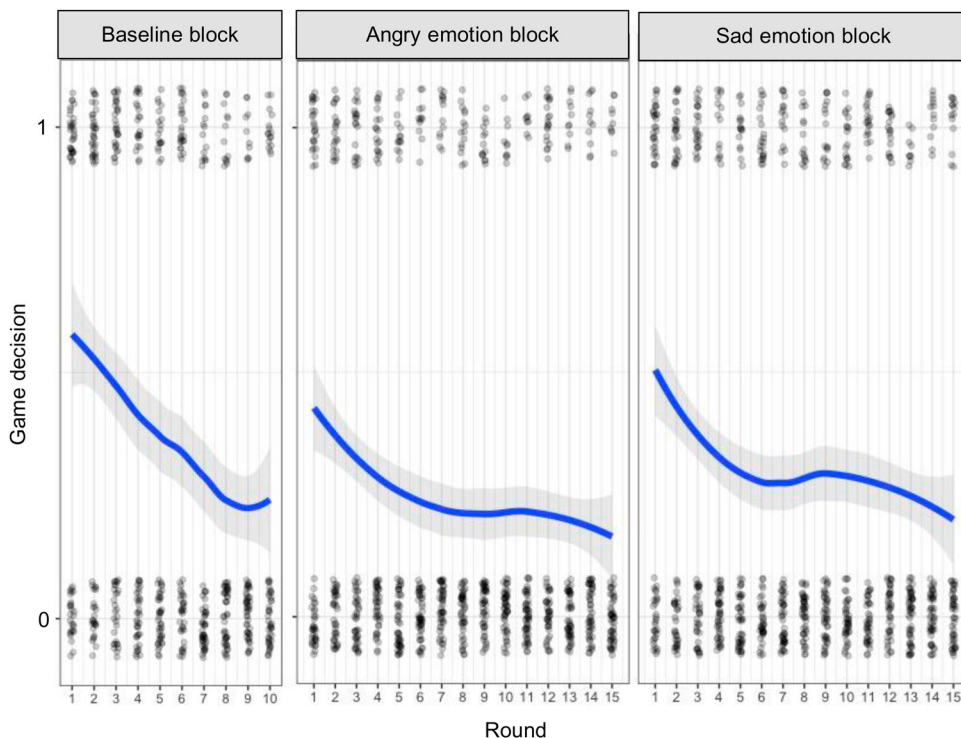
## Results

We carried out the preregistered analyses when 60 participants (mean age = 24.8; 39 females, 17 males, and 4 non-binary) were recruited as per our preregistered sequential analysis plan. Among this sample, 51.67% of participants were White; 38.33% were Asian or Asian British; 1.67% were Black, African,

Black British or Caribbean; 1.67% belonged to mixed or multiple ethnic groups; 5% were from other ethnic groups; and 1.67% preferred not to report. Considering the COVID-related restrictions on in-person testing at University of Glasgow and Macquarie University between September and December 2021, all data were collected at the University of Glasgow. Therefore, the exploratory model to control for the potential random effects induced by collecting data at two sites was not performed. We measured participants' daily exposure to robots (Riek et al., 2011) to ensure that they were generally naïve to robots. In the question of how many robot-related films participants had seen before (from a list of 14 films including *Westworld*, *Real Humans*, etc), the median number of robot films seen was 3, with an interquartile range (IQR) of 3. When asking participants how often they engaged with robots in their daily life on a scale from 1 (Never) to 7 (Daily), the median response was 2 (IQR=2). The results confirmed that participants did not have extensive experience with robots before taking part in this study, and therefore their a priori understanding of robots was unlikely to impact the current HRI.

First, we visualised the distribution of participants' binomial game decisions (to share coins with the robots or not) in the three blocks in Figure 8. From Figure 8, we could see that the cooperative trends of the three game blocks were similar. Participants started from a higher cooperative tendency in the beginning of each block, and this tendency decreased until the end of the game. The only visible difference between the baseline block and the two emotion blocks was that participants were making slightly more cooperative choices near the end of the block. However, since we did not inform participants of the total number of rounds for each block, it was unlikely that the increasing cooperative decisions were planned deliberately by participants.

Second, we calculated the mean cooperation rates for each block by dividing the numbers of participants' cooperative decisions by the total numbers of game rounds (10 rounds in the baseline block and 15 rounds for each emotion block). In the baseline block, the mean cooperation rate was 37.13%; in the angry block it was 24.83%; in the sad block it was 30.34%. Following the registered analysis plan, we reported the main result of a linear mixed effects



**Figure 8** . Binomial game decision distribution across the three game blocks (sharing coded as 1; keeping coded as 0). Nonparametric smoothed curves were added to show the cooperative trends.

model to examine whether there was an interplay between cooperative predisposition and the robots' emotions. For exploratory analyses, we presented the results of the registered model which included an additional factor of participants' emotion perception accuracy. Additionally, we conducted and reported the results of unregistered exploratory analyses, which were the logistic version of the registered models. The logistic models used participants' binomial decisions as the dependent variable, instead of the log-transformed cooperation rates. We carried out this additional modelling because we realised the process of log-transformation (in order to feed the data of cooperation rates to linear models) led to information loss, while using mixed effects logistic regression models on the raw dataset might bring about higher power to detect the effects of interest. Below we present each part of these analyses in detail.

### Main model results

The model successfully converged with the pre-registered model design. We included the fixed factors of the robots' emotions (anger and sadness) and participants' cooperative predisposition (i.e. log-transformed cooperation rates in the baseline block), the dependent variable of the log-transformed cooperation rates in the two emotion blocks, and the random effects of subject-level random intercepts. As mentioned above, we adopted sequential analyses (with two interim analyses) and therefore we used  $p = .0221$  as the adjusted alpha level (Pocock, 1977). We found a significant factor of participants' cooperative predisposition in this model ( $\beta = 0.54$ , 95% CI [0.17, 0.92],  $p = .004$ ,  $\eta_p^2 = .23$ ). However, neither the fixed effect of the robots' emotions ( $\beta = 0.34$ , 95% CI [-0.01, 0.69],  $p = .058$ ,  $\eta_p^2 = .07$ ) nor the interaction between the two factors ( $\beta = 0.06$ , 95% CI [-0.41, 0.53],  $p = .795$ ,  $\eta_p^2 = .001$ ) was significant. Based on our registered sampling plan of sequential analyses, the data collection was stopped given that the effect size (Cohen's  $f^2 = 0.0004$ ) of the interaction (the main effect of interest) is smaller than the SESOI ( $f^2 = 0.02$ ). Namely, the true effect size of the interaction might be smaller than what was considered to be practically meaningful. Therefore, we decided not to pursue such a minor effect with a bigger sample size. Overall, the  $R^2$  of the model was .330, with the fixed effects  $R^2 = .178$  and the random effects  $R^2 = .153$ .

### Registered exploratory model results

In the registered exploratory model, we included an additional fixed factor — the binomial records of whether participants had accurately perceived the robots' emotion as we expected — into the design of the main model. The answers we coded as “successfully perceived the robot's anger” included participants' reports of “angry”, “anger”, “furious” that were used to describe the robot programmed to display anger; the answers we coded as “successfully perceived the robot's sadness” were the reports that explicitly used the words of “sad” or “sadness” to describe the robot programmed to display sadness. Since the manipulation check was measure by open-ended questions and we did not provide any word bank for participants to choose from, a few participants would use the words that were more ambiguous, like “disappointed”, “frustrated”, “discontent”, “displeasure”, to describe the robots' emotional displays. We did not include those answers as evidence of successfully perceiving the emotional manipulation. Also, three participants reported perceiving both negative emotions in a single emotion block: two said they perceived both sadness and anger from the robot programmed to display sad expressions, and one perceived both anger and sadness from the robot programmed to display angry expressions. We also excluded these reports from correct emotional recognition. All in all, the successful perception rate for the robot's angry display was 66.7%, and the rate for the sad display was 51.7%.

We then added this binomial variable of whether participants perceived the robots' emotional manipulation into the model, to examine the extent to which individual differences in emotion perception might influence the results. The model output was presented in Table 1. We found that none of the fixed factors, nor their interactions, significantly impacted people's cooperative tendencies.

Overall, the  $R^2$  of the registered exploratory model was .347, with the fixed effects  $R^2 = .187$  and the random effects  $R^2 = .160$ . We conducted a model comparison test by the R function `anova()` to examine whether inclusion of the additional factor (“*subj\_perception*”) improved the model fit. The result suggested that the difference between the main model (without the “*subj\_perception*” factor) and the registered exploratory model (with the “*subj\_perception*” factor) was not significant,  $\chi^2(4, 106) = 1.72$ ,  $p = .79$ .



**Table 1.** Results of the linear mixed effects model that examined the effects of the robots' emotions, participants' cooperative predisposition, and their emotion perception accuracy on subjects' log-transformed cooperation rates

	Registered exploratory model					
	<i>cooperation ~ emotion*coop_predisposition*subj_perception + (1   subj_id)</i>					
	<i>Estimate</i>	<i>SE</i>	<i>Low CI</i>	<i>High CI</i>	$\eta_p^2$	<i>p-value</i>
intercept	-0.88	0.28	-1.43	-0.33		<b>.002*</b>
emotion [sad-angry]	0.38	0.33	-0.27	1.03	.06	.250
coop_predisposition	-0.02	0.53	-1.04	1.01	.08	.977
subj_perception [correct-incorrect]	0.04	0.32	-0.60	0.67	.00005	.914
emotion* coop_predisposition	0.63	0.57	-0.48	1.74	.01	.265
emotion * subj_perception	-0.04	0.43	-0.89	0.81	.00009	.926
coop_predisposition* subj_perception	0.63	0.56	-0.47	1.73	.008	.264
emotion* coop_predisposition* subj_perception	-0.65	0.66	-1.95	0.65	.01	.326
Subject-level random intercepts	0.39					
Residuals	0.75					

*CI* = 95% confidence interval.

\* $p < .0221$

Abbreviations: *SE* = standard error; *CI* = confidence interval.

### Unregistered exploratory model results

Although the usage of log-transformed cooperation rates allowed the dependent variable values to range from negative to positive, instead of 0–1 (Benoit, 2011), we lost some data points because if a participant made no cooperative decision in a game block, the 0 cooperation rate would lead to negative infinity after being log-transformed. This resulted in us having to exclude 14 data points (which resulted in this negative infinity value after log transformation) in order to run linear mixed effects models. Excluding these data points caused crucial information loss since those data represented performances by the most competitive individuals. Therefore, we conducted additional mixed effects logistic regression models to examine if the effects of interest would be better to detect by performing analyses on the raw and complete dataset (binomial game decisions: cooperative decisions coded as 1 and noncooperative decisions coded as 0).

First, in the logistic version of the main model (Model 1 in Tables 2 and 3), we used participants' binomial game decisions as the dependent variable and added the random intercepts of game rounds into the random effect structure. The rest of the model

design remained the same as the main linear model. Similar to the results of the main model, we found a significant effect from participants' cooperative predisposition ( $\beta = 3.71$ , 95% *CI* [2.16, 5.26],  $p < .001$ ) whereas the main effect of the robots' emotions ( $\beta = 0.25$ , 95% *CI* [-0.41, 0.92],  $p = .452$ ) and the interaction between the two factors ( $\beta = 0.15$ , 95% *CI* [-1.39, 1.69],  $p = .851$ ) were nonsignificant.

Second, we ran a logistic version of the registered exploratory model which included the factor of individuals' emotion perception. Again, we controlled for the round-level random effects in the logistic models. We started with the most complex random structure (Barr et al., 2013) for round-level random effects – ( $1 + emotion*subj\_perception | round$ ) – but the model failed to converge and we therefore run the model with only the random intercepts of subjects (Model 2 in Table 2). Results yielded a significant effect from subjects' emotion perception accuracy ( $\beta = -1.74$ , 95% *CI* [-3.15, -0.33],  $p = .015$ ) whereas all other fixed effects and their interaction were nonsignificant (Model 2 in Table 3). In general, people who correctly perceived the robots' angry and sad emotions were less likely to cooperate with the robots in emotion blocks. Given the complexity of

**Table 2.** The designs of the three unregistered exploratory models to examine the effects of the robots' emotions, participants' cooperative predisposition, and their emotion perception accuracy on subjects' binomial game decisions

	Model design		
	<i>Fixed factor(s)</i>	<i>Random effects</i>	<i>Dependent variable</i>
Model 1	<i>emotion*coop_predisposition</i>	(1   <i>subj_id</i> ) + (1   <i>round</i> )	<i>game decisions</i>
Model 2	<i>emotion*coop_predisposition*subj_perception</i>	(1   <i>subj_id</i> )	<i>game decisions</i>
Model 3	<i>coop_predisposition*subj_perception</i>	(1   <i>subj_id</i> ) + (1   <i>round</i> )	<i>game decisions</i>

the three factors involved in the model, we visualised the overall results of the Model 2 in Figure 9 by the R package “effects” (v4.1.4) (Fox & Weisberg, 2018). From Figure 9, it is possible to see a positive correlation between people’s cooperative tendencies in the baseline block and their cooperative probability in emotion blocks, and the correlation might be shaped by people’s emotion perception accuracy (albeit the interaction was not significant  $p = .059$ , by the alpha level of  $p = .0221$ ).

Finally, for exploratory purposes, we ran the Model 3 without the factor of the robots’ emotions since its effect did not seem significant in either Model 1 or Model 2. In the result of Model 3, the effect of people’s cooperative predisposition became significant ( $\beta = 3.05$ , 95% CI [1.23, 4.87],  $p = .001$ ), and the effect of subjects’ emotion perception accuracy was not significant ( $\beta = -0.70$ , 95% CI [-1.58, 0.17],  $p = .115$ ) given the pre-defined alpha level of .0221. The output summary of three models and the result of model comparison are reported in Table 3. Among the three logistic models, none of these three models showed significant improvement in model fit compared to the other two models.

## Discussion

In this study, we sought to examine the extent to which people’s cooperative tendencies in prisoner’s dilemma (PD) games are influenced by robots’ negative emotion displays and whether the influence of robotic emotion displays is shaped by individual

participants’ cooperative predispositions (measured in a baseline game block where the robot did not display any emotion). Based on Van Kleef et al.’s (2010) Emotion as Social Information (EASI) model, we predicted that participants who were more cooperative in the baseline block would become even more cooperative when the robot displayed sadness (to show compassion), but less cooperative when the robot displayed anger (to punish who eroded cooperative atmosphere), whereas participants who were competitive in the baseline block would be made to cooperate by the robot’s anger (to avoid lose-lose dispute) and would be even more competitive by the robot’s sadness (to take advantage of the signs of weakness). The first interim analysis carried out when 60 participants were recruited failed to support these predictions. What has emerged is a significant effect of people’s cooperative predispositions on their cooperative tendencies towards both emotional robots. Based on our preregistered sequential analysis plan, we did not continue further data collection given that the effect size of the main effect of interest (the interaction between the robots’ emotions and people’s cooperative predisposition) was smaller than the pre-defined SESOI. Below we discuss our findings in detail.

We performed a linear mixed effects model to examine the main research question and expected to find a significant interaction between the robots’ emotions and participants cooperative predispositions on their log-transformed cooperation rates in emotion blocks. However, only the main effect of people’s

**Table 3.** The result summary of the three unregistered exploratory models and the outcome of the model comparison.

	Unregistered exploratory models					
	Model 1		Model2		Model 3	
	Estimate (SE)	<i>p</i>	Estimate (SE)	<i>p</i>	Estimate (SE)	<i>p</i>
intercept	-2.68 (0.35)	<.001*	-1.20 (0.64)	.061	-2.03 (0.42)	<.001*
emotion [sad-angry]	0.25 (0.34)	.452	-1.07 (0.71)	.129		
coop_predisposition	3.71 (0.79)	<.001*	1.09 (1.47)	.460	3.05 (0.93)	.001*
subj_perception [correct-incorrect]			-1.74 (0.72)	.015*	-0.70 (0.45)	.115
emotion* coop_predisposition	0.15 (0.79)	.851	2.52 (1.59)	.113		
emotion * subj_perception			1.55 (0.88)	.077		
coop_predisposition* subj_perception			3.09 (1.64)	.059	0.95 (1.00)	.342
emotion* coop_predisposition* subj_perception			-2.71 (2.04)	.183		
df			3		0	
AIC	1973		1998		1976	
BIC	2006		2048		2009	
Log-likelihood	-980		-990		-982	
$\chi^2$			0.00		0.00	
<i>p</i>			1		1	

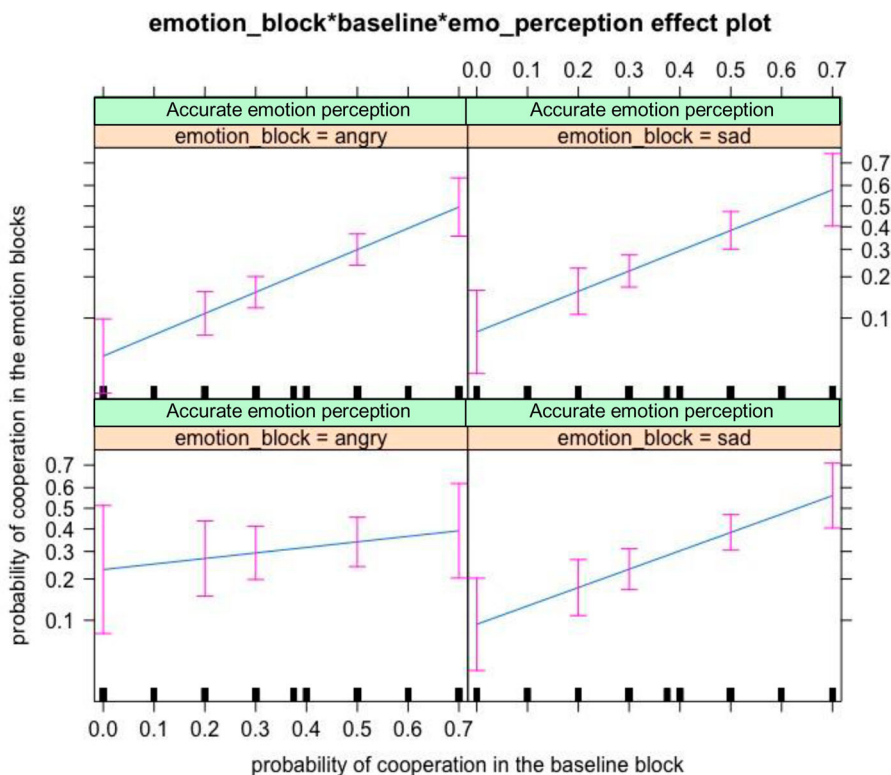
CI = 95% confidence interval.

\* $p < .0221$

Abbreviations: SE = standard error.

cooperative predisposition was found to be significant with a large effect size. Participants who showed stronger cooperative tendencies in the baseline block were more likely to cooperate with the robots in emotion blocks. The effect of cooperative predisposition was confirmed by the logistic version of the main model, which used people's binomial decisions as the dependent variable instead of the log-transformed data. This high behavioural consistency within individuals might imply that participants' game decisions in the baseline block reflected their innate cooperative attitudes in this context. Previous research has pointed out the concept of Social Value Orientation (SVO), which refers to people's dispositional prosocial tendencies during interpersonal interactions (Murphy & Ackermann, 2014). The impact of SVO on cooperative decisions when faced with a social dilemma has been confirmed by at least two meta-analysis studies (Balliet et al., 2009; Pletzer et al., 2018). This work has shown a consistent medium effect of SVO on social decisions. In the present study, we did not include a self-report SVO measure (e.g. the scale by Murphy

et al., 2011), because our previous work addressing related questions (Hsieh et al., 2020), did not provide any evidence for a significant relationship between participants' SVO scores and their cooperative decisions during PD games played with a Cozmo robot, and almost all participants were categorised to the "prosocial" SVO type. It is consequently worth questioning to what extent people's self-reports of SVO are influenced by social desirability and whether there is a link between the SVO measure (which was not specifically designed to measure the attitudes towards robots) and people's actual cooperative behaviours in HRI. Although we cannot say for sure if participants' consistent cooperative tendencies throughout the three game blocks were associated with SVO, our current finding highlights the strong effects of personal factors in cooperation with robots. Furthermore, it seems such top-down personal effects might surpass the bottom-up emotional displays presented by the robots in our experiment. Also, this finding confirms the utility of the baseline measure. Even though our baseline block only involved 10 game rounds



**Figure 9.** Effect plot of the unregistered Model 2. The model examined the effects of the robots' emotions, participants' cooperative predisposition and emotion perception accuracy on participants' binomial decisions in the PD games.

(compare to 15 rounds in each emotion block), participants' cooperation rates were still predictive as to what they would do in similar scenarios.

However, we were surprised to find that participants' cooperative decisions in the final two rounds of the baseline block seemed to increase a little, reversing the decline in cooperation rates that was observed in previous rounds of the baseline block and in both emotion blocks (Figure 8). One possible reason behind this could be the robot's reciprocal (tit-for-tat) game strategy adopted in the second half of the baseline block. A previous study has shown that a robot's tit-for-tat strategy, compared to a random strategy, in PD games led to higher cooperation rates among participants (Sandoval et al., 2016). We programmed our three robot players to always start with a fixed sequence of decisions, followed by a tit-for-tat strategy, across all the game blocks, in order to make their game strategies less predictable and to increase the chances of exposing participants to the robot emotion manipulation. Still, it was possible that near the end of the baseline block, participants realised the robot's tit-for-tat strategy, especially when the robot did not display any emotional reaction to distract them, and therefore became more willing to cooperate. However, this interpretation remains speculative at this stage, and we further research will be required to substantiate this explanation. Currently, we cannot exclude the possibility that this finding was simply due to random variance within our sample.

In light of the well-documented individual differences in emotion perception of human facial expressions (Barrett et al., 2019) and of robots' emotion displays (Stock-Homburg, 2022), we planned to explore if the variation in emotion perception would influence participants' cooperative tendencies in PD games and the effects of the robots' emotions. In participants' self-report data concerning observed emotions from the two emotional robots, we did find considerable individual differences in perceiving and reporting the robots' emotional displays. Although more than half of the participants correctly recognised that one of the robots showed sad expressions and the other was angry, some participants described them only in comparative terms (e.g. saying one robot was less angry than the other) or were not aware of any emotional displays by the robots. Quite a few participants seemed to perceive and describe only the negativity of the emotions displayed by the robots and reported the expressions as "displeasure", "frustration",

or "disappointment", without explicitly identifying them as sadness or anger. The result of the accuracy rates in perceiving the robots' sadness and anger suggested that the robot's angry expression was easier for participants to recognise, which verifies the conclusion of Stock-Homburg's (2022) review paper suggesting that robots' higher arousal emotions (e.g. anger and happiness) are more consistently and accurately perceived by people (Stock-Homburg, 2022). In the review paper, Stock-Homburg (2022) extensively reviewed 43 studies that examined the emotional expressions displayed by (1) the robots that only have robotic faces (e.g. Barthoc robot, EMYS robot); (2) the robots with anthropomorphic full bodies (e.g. NAO, Pepper robot); and (3) zoomorphic robots (e.g. Keepon robot, KAROTZ robot). Our findings of Cozmo robots therefore added another example of non-humanlike robots whose high-arousal emotional displays are better recognised by people.

To statistically examine the impact of individual differences in perceiving robots' emotional displays, we ran both linear and logistic mixed effects models. We found a significant effect of emotion perception only in the logistic model with all the factors – including the robots' emotions, people's cooperative predisposition and individual emotion perception – involved (Model 2 in Table 3). Participants who correctly perceived the robots' negative emotions displayed after being betrayed by a human player in PD games were less likely to cooperate with the robots in PD games. However, the effect was not significantly shaped by the robot's emotion types (sadness or anger), nor by people's cooperative predisposition, against our predictions. In the current study, the effect of the robots' emotional displays might be constrained by the low recognition rates for robotic emotions in the embodied human–robot PD games (66.7% accuracy for anger; 51.7% for sadness), which were much lower than the recognition rates we measured in our online pilot (98.4% accuracy for anger; 90.6% for sadness). When engaging in economic games played with embodied robots, people might attend mostly to strategic decision-making in order to win, and have limited attention paid to the robot opponents' emotional expressions during games. Although we manipulated the robots so that their emotional displays occurred after each round, when participants were not required to make any other game response, it is still possible that participants were more focused on their next step in the game, and therefore were not fully aware (or focussed on) what the robots were doing.

Contrarily, when examining the influence of individual emotion perception via a linear mixed effects model on the log-transformed dependent variable, we did not find any significant effect from the fixed factors and their interactions. We think these results can be explained by the fact that, when running the linear model, we excluded 14 data points to fix the issue of zero cooperation rates leading to values of negative infinity. This data exclusion also meant we lost performance data from the most competitive participants. Therefore, the usage of mixed effects logistic regression models gave us more power to detect the effects of interest, and brought about more complete results since the analyses were performed on the entire dataset. The reason why we did not plan on logistic models in the first place was due to the difficulty in performing beta weight estimation for power analyses given the limited number of studies adopting logistic mixed effects model approach in the literature. One study by Moisan et al. (2018) that used this statistical approach focused on the effects of incentive structures on cooperation in interpersonal PD games, rather than robots' emotional displays in human-robot PD games. Consequently, we suggest that more research could consider using mixed effects logistic regression models for analysing such binomial decision data. The strengths of mixed effects models to control for subject-level and stimulus-level random variation also make them outperform ANOVAs or t-tests in many cases (Debruine & Barr, 2019; Field & Wright, 2011).

Among the three exploratory logistic models we conducted, only the personal factors (including cooperative predisposition and individual emotion perception) were found to be relevant to people's cooperative tendencies towards the robots in PD games. Individual differences in emotion perception and cooperative predisposition, compared to the robots' emotion displays, seemed to play a more important role in explaining people's cooperative decisions in the current human-robot PD games. Similar to our finding in the main model, the personal factors drove participants' game decisions more than the robots' emotion types did. Kjell and Thompson's (2013) study also demonstrated the power of personal factors in social decision-making process and found that individuals' SVO outweighed the influence of the essay emotion manipulation tasks on the subjects' cooperative decisions in a computer-mediated PD game. However, since the emotion recognition rates for Cozmo's sad and angry displays were lower than

our expectations in this current study, we are unable to state decisively whether personal factors are more relevant than robots' emotional displays to people's cooperative willingness during HRIs in general. Follow-up studies are warranted for a more robust understanding of the effects of robots' emotional displays on people's cooperative decisions in embodied HRIs, and for clarifying how the effects of robotic emotions relate to personal factors, such as cooperative predispositions and emotion perception. Future research could consider adopting less cognitive demanding game scenarios to examine the effects of robotic emotional displays on people's cooperative tendencies, in order to ensure participants have the cognitive resources available to process robots' emotional displays (and other responses) while engaging in social decision-making tasks.

So far, we cannot reject the null hypothesis and cannot claim that people's cooperative decisions in the human-robot PD games are influenced by the interaction between the robots' emotions displays (anger and sadness) and people's cooperative predisposition in the way as the EASI model proposed (Van Kleef et al., 2010). However, it is important to emphasise that the EASI model was derived from human psychological literature and was originally intended to explain and predict interpersonal effects of emotional cues during interpersonal interactions between two people. Therefore, the EASI model might not be the most suitable model to predict the impact of embodied robots' emotional displays on people's cooperative decisions. This also demonstrates the limitations of understanding HRIs merely through the lens of human social cognition, while disregarding the fact that social robots may be seen or categorised variably across a continuum that ranges from simple inanimate objects through to humans, given the vast variety in robots' physical features and social characteristics (Cross & Ramsey, 2021). As such, a robot-specific theoretical framework would be helpful if we are to better explain and predict the social effects of artificial agents' emotional displays on people's behaviours.

Moreover, other factors are also likely to influence people's cooperative tendencies towards robots that were not adequately captured in this study, such as individuals' intergroup perceptions towards robots (De Jong et al., 2021; Fraune et al., 2017), anthropomorphism (Torta et al., 2013), trust towards robots (Paeng et al., 2016; Tulk & Wiese, 2018; Wu et al., 2016) and the type of game strategy adopted by robot opponents (de Melo & Terada, 2020). In this

study, we focused exclusively on the effects of Cozmo robots' sad and angry displays, while attempting to control for other individual random variation via mixed effects modelling. Future studies have the opportunity to expand the present investigation by examining the social effects of other robotic emotional displays, since current evidence has shown that virtual agents' joy and regret expressions might be particularly impactful on people's cooperative tendencies, compared to displays of sadness and anger (de Melo, Carnevale, et al., 2014; de Melo & Terada, 2019, 2020). Also, follow-up studies could further investigate additional personal, robotic, and contextual factors in PD games for an in-depth and comprehensive understanding of the decision-making process in human–robot cooperation.

Nevertheless, the present findings underscore the utility and importance of performing a manipulation check for emotion manipulation on robots and deploying a baseline measure for people's dispositional cooperative tendencies. Especially for between-subject design or small sample size studies, it is essential to ensure that people's cooperative decisions are driven by the experimental manipulation, rather than by their innate cooperative tendencies or by individual differences in perception. Also, when investigating the social effects of embodied robotic emotions, it is worth conducting pilot studies in more realistic scenarios where people are observing real-life, embodied HRIs, rather than simply checking stimulus validity via online experiments (c.f., Cross & Ramsey, 2021; Henschel et al., 2020). It could be the case that the actual effectiveness of emotional manipulation on robots is overestimated in complex and dynamic embodied HRIs. By taking these considerations into account, researchers could truly reveal the potential effects of robots' emotional displays on shaping people's cooperative decisions.

## Acknowledgements

We thank Kohinoor Darda and Nathan Caruana for the feedback on the manuscript submitted for stage 1 registered report review and Bishakha Chaudhury and Amol Deshmukh for the technical support. The work was supported by the European Research Council under the European Union's 2020 research and innovation program (Grant agreement 677270 to ESC) and Leverhulme Trust (PLP-2018-152 to ESC).

## Disclosure statement

No potential conflict of interest was reported by the author(s).

## Funding

This work was supported by H2020 European Research Council: [Grant Number ERC-StG-2015-677270]; The Leverhulme Trust: [Grant Number PLP-2018-152].

## Author Contributions

**TYH:** Conceptualization, Methodology, Investigation, Data Analysis and Curation, Writing, Visualization;  
**ESC:** Conceptualization, Writing, Supervision

## ORCID

Te-Yi Hsieh  <http://orcid.org/0000-0002-4746-9303>

Emily S. Cross  <http://orcid.org/0000-0002-1671-5698>

## Reference

- Adam, H., & Brett, J. M. (2015). Context matters: The social effects of anger in cooperative, balanced, and competitive negotiation situations. *Journal of Experimental Social Psychology*, 61, 44–58. <https://doi.org/10.1016/j.jesp.2015.07.001>
- Anvari, F., & Lakens, D. (2019). *Using anchor-based methods to determine the smallest effect size of interest*. [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/syp5a>.
- Arslan, R. C., Walther, M. P., & Tata, C. S. (2020). Formr: A study framework allowing for automated feedback generation and complex longitudinal experience-sampling studies using R. *Behavior Research Methods*, 52(1), 376–387. <https://doi.org/10.3758/s13428-019-01236-y>
- Axelrod, R., & Hamilton, W. D. (1981). The Evolution of cooperation. *Science*, 211(4489), 1390–1396. <http://doi.org/10.1126/science.7466396>
- Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes and Intergroup Relations*, 12(4), 533–547. <https://doi.org/10.1177/1368430209105040>
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 255–278. <https://doi.org/10.1016/j.jml.2012.11.001>
- Barratt, D., Rédei, A. C., Innes-Ker, Å, & van de Weijer, J. (2016). Does the Kuleshov effect really exist? Revisiting a classic film experiment on facial expressions and emotional contexts. *Perception*, 45(8), 847–874. <https://doi.org/10.1177/03010066166638595>
- Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional expressions reconsidered: Challenges to inferring emotion from human facial movements. *Psychological Science in the Public Interest*, 20(1), 1–68. <https://doi.org/10.1177/1529100619832930>
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects: Models using lme4. *Journal of Statistical Software*, 67(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>

- Benoit, K. (2011). Linear regression models with logarithmic transformations. *London School of Economics*, 22(1), 23–36. <https://kenbenoit.net/assets/courses/ME104/logmodels2.pdf>
- Bland, A. R., Roiser, J. P., Mehta, M. A., Schei, T., Sahakian, B. J., Robbins, T. W., & Elliott, R. (2017). Cooperative behavior in the ultimatum game and prisoner's dilemma depends on players' contributions. *Frontiers in Psychology*, 8, 1–11. <https://doi.org/10.3389/fpsyg.2017.01017>
- Bonnefon, J.-F., Shariff, A., & Rahwan, I. (2016). The social dilemma of autonomous vehicles. *Science*, 352(6293), 1573–1576. <https://doi.org/10.1126/science.aaf2654>
- Bravo, G., Squazzoni, F., & Boero, R. (2012). Trust and partner selection in social networks: An experimentally grounded model. *Social Networks*, 34(4), 481–492. <https://doi.org/10.1016/j.socnet.2012.03.001>
- Broadbent, E. (2017). Interactions with robots: The truths We reveal about ourselves. *Annual Review of Psychology*, 68, 627–652. <https://doi.org/10.1146/annurev-psych-010416-043958>
- Cahapay, M. B. (2020). Rethinking education in the New normal post-COVID-19 Era: A curriculum studies perspective. *Aquademia*, 4(2), 1–5. <https://doi.org/10.29333/aquademia/8315>
- Capraro, V., Jagfeld, G., Klein, R., Mul, M., & de Pol, I. v. (2019). Increasing altruistic and cooperative behaviour with simple moral nudges. *Scientific Reports*, 9(1), 1–11. <https://doi.org/10.1038/s41598-019-48094-4>
- Capraro, V., & Perc, M. (2021). Mathematical foundations of moral preferences. *Journal of The Royal Society Interface*, 18(175), 1–13. <https://doi.org/10.1098/rsif.2020.0880>
- Chaudhuri, A., Sopher, B., & Strand, P. (2002). Cooperation in social dilemmas, trust and reciprocity. *Journal of Economic Psychology*, 23(2), 231–249. [https://doi.org/10.1016/S0167-4870\(02\)00065-X](https://doi.org/10.1016/S0167-4870(02)00065-X)
- Chaudhury, B., Hortensius, R., Hoffmann, M., & Cross, E. S. (2020). *Tracking human interactions with a commercially-available robot over multiple days: A tutorial* [Preprint]. PsyArXiv. <https://doi.org/10.31234/osf.io/fd3h2>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates.
- Cross, E. S., & 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>
- Cross, E. S., Riddoch, K. A., Pratts, J., Titone, S., Chaudhury, B., & Hortensius, R. (2019). A neurocognitive investigation of the impact of socializing with a robot on empathy for pain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 374(1771), 20180034. <https://doi.org/10.1098/rstb.2018.0034>
- Darwin, C., & Prodger, P. (1998). *The expression of the emotions in man and animals*. Oxford University Press.
- Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1480), 679–704. <https://doi.org/10.1098/rstb.2006.2004>
- Debruine, L. M., & Barr, D. J. (2019). Understanding mixed effects models through data simulation. *PsyArXiv*.
- De Jong, D., Hortensius, R., Hsieh, T.-Y., & Cross, E. S. (2021). Empathy and schadenfreude in human-robot teams. *Journal of Cognition*, 4(1), 35. <https://doi.org/10.5334/joc.177>
- de Melo, C. M., 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](https://doi.org/10.1007/978-3-642-15892-6_38)
- de Melo, C. M., Carnevale, P., & Gratch, J. (2011). The effect of expression of anger and happiness in Computer agents on Negotiations with humans. *10th International Conference on Autonomous Agents and Multiagent Systems AAMAS 2011, Aamas*, 937–944. <https://doi.org/10.1016/j.jclepro.2016.12.062>
- de Melo, C. M., Carnevale, P. J., Read, S. J., & Gratch, J. (2014). Reading people's minds from emotion expressions in interdependent decision making. *Journal of Personality and Social Psychology*, 106(1), 73–88. <https://doi.org/10.1037/a0034251>
- de Melo, C. M., Gratch, J., & Carnevale, P. J. (2014). Humans vs. Computers: Impact of emotion expressions on people's decision making. *IEEE Transactions on Affective Computing*, 1(2), 1–11. <https://doi.org/10.1109/TAFFC.2014.2332471>
- de Melo, C. M., Marsella, S., & Gratch, J. (2019). Human cooperation when acting through autonomous machines. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.1817656116>
- de Melo, C. M., & Terada, K. (2019). Cooperation with autonomous machines through culture and emotion. *PLOS ONE*, 14(11), e0224758. <https://doi.org/10.1371/journal.pone.0224758>
- de Melo, C. M., & Terada, K. (2020). The interplay of emotion expressions and strategy in promoting cooperation in the iterated prisoner's dilemma. *Scientific Reports*, 10(1), 1–8. <https://doi.org/10.1038/s41598-020-71919-6>
- Elliot, A. J., Jury, M., & Murayama, K. (2018). Trait and perceived environmental competitiveness in achievement situations. *Journal of Personality*, 86(3), 353–367. <https://doi.org/10.1111/jopy.12320>
- Embrey, M., Fréchette, G. R., & Yuksel, S. (2018). Cooperation in the finitely repeated prisoner's dilemma\*. *The Quarterly Journal of Economics*, 133(1), 509–551. <https://doi.org/10.1093/qje/qjx033>
- Field, A. P., & Wright, D. B. (2011). A primer on using multilevel models in clinical and experimental Psychopathology research. *Journal of Experimental Psychopathology*, 2(2), 271–293. <https://doi.org/10.5127/jep.013711>
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8(15), 27. <https://doi.org/10.18637/jss.v008.i15>
- Fox, J., & Weisberg, S. (2018). Visualizing Fit and lack of Fit in complex regression models with predictor effect plots and partial residuals. *Journal of Statistical Software*, 87(9), 1–27. <https://doi.org/10.18637/jss.v087.i09>
- Franken, I. H. A., & Muris, P. (2005). Individual differences in decision-making. *Personality and Individual Differences*, 39(5), 991–998. <https://doi.org/10.1016/j.paid.2005.04.004>
- Fraune, M. R., Šabanović, S., & Smith, E. R. (2017). Teammates first: Favoring ingroup robots over outgroup humans. *2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*, 1432–1437. <https://doi.org/10.1109/ROMAN.2017.8172492>
- Frijda, N. H. (1986). *The emotions*. Cambridge University Press.

- Galak, J., LeBoeuf, R. A., Nelson, L. D., & Simmons, J. P. (2012). Correcting the past: Failures to replicate psi. *Journal of Personality and Social Psychology*, 103(6), 933–948. <https://doi.org/10.1037/a0029709>
- George, J. M., & Dane, E. (2016). Affect, emotion, and decision making. *Organizational Behavior and Human Decision Processes*, 136, 47–55. <https://doi.org/10.1016/j.obhdp.2016.06.004>
- Green, P., & Macleod, C. J. (2016). Simr: An R package for power analysis of generalized linear mixed models by simulation. *Methods in Ecology and Evolution*, 7(4), 493–498. <https://doi.org/10.1111/2041-210X.12504>
- Grossman, R. B., Zane, E., Mertens, J., & Mitchell, T. (2019). Facetime vs. Screentime: Gaze patterns to live and video social stimuli in adolescents with ASD. *Scientific Reports*, 9(1), 12643. <https://doi.org/10.1038/s41598-019-49039-7>
- Hamann, S., & Canli, T. (2004). Individual differences in emotion processing. *Current Opinion in Neurobiology*, 14(2), 233–238. <https://doi.org/10.1016/j.conb.2004.03.010>
- Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the Age of human–robot interaction. *Trends in Neurosciences*, S0166223620300734. <https://doi.org/10.1016/j.tins.2020.03.013>
- 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. *Proceedings of the 18th International Conference on Intelligent Virtual Agents*, 275–280. <https://doi.org/10.1145/3267851.3267911>
- Hortensius, R., & Cross, E. S. (2018). From automata to animate beings: The scope and limits of attributing socialness to artificial agents: Socialness attribution and artificial agents. *Annals of the New York Academy of Sciences*, 1426(1), 93–110. <https://doi.org/10.1111/nyas.13727>
- Hortensius, R., Hekele, F., & Cross, E. S. (2018). The perception of emotion in artificial agents. *IEEE Transactions on Cognitive and Developmental Systems*, 10(4), 852–864. <https://doi.org/10.1109/TCDS.2018.2826921>
- Hsieh, T.-Y., Chaudhury, B., & Cross, E. S. (2020). Human-robot cooperation in economic games: People show strong reciprocity but conditional prosociality toward robots [preprint]. *PsyArXiv*, 1–34. <https://doi.org/10.31234/osf.io/q6pv7>
- Ito, H., & Tanimoto, J. (2018). Scaling the phase-planes of social dilemma strengths shows game-class changes in the five rules governing the evolution of cooperation. *Royal Society Open Science*, 5(10), 181085. <https://doi.org/10.1098/rsos.181085>
- Jamaludin, S., Azmir, N. A., Mohamad Ayob, A. F., & Zainal, N. (2020). COVID-19 exit strategy: Transitioning towards a new normal. *Annals of Medicine and Surgery*, 59, 165–170. <https://doi.org/10.1016/j.amsu.2020.09.046>
- Janssen, M. A. (2008). Evolution of cooperation in a one-shot prisoner's dilemma based on recognition of trustworthy and untrustworthy agents. *Journal of Economic Behavior & Organization*, 65(3–4), 458–471. <https://doi.org/10.1016/j.jebo.2006.02.004>
- Kayukawa, Y., Takahashi, Y., Tsujimoto, T., Terada, K., & Inoue, H. (2017). Influence of emotional expression of real humanoid robot to human decision-making. *2017 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 1–6. <https://doi.org/10.1109/FUZZ-IEEE.2017.8015598>
- Kim, S.(Sam), Kim, J., Badu-Baiden, F., Giroux, M., & Choi, Y. (2021). Preference for robot service or human service in hotels? Impacts of the COVID-19 pandemic. *International Journal of Hospitality Management*, 93, 102795. <https://doi.org/10.1016/j.ijhm.2020.102795>
- Kjell, O. N. E., & Thompson, S. (2013). Exploring the impact of positive and negative emotions on cooperative behaviour in a prisoner's dilemma game. *PeerJ*, 1(2000), e231. <https://doi.org/10.7717/peerj.231>
- Kopelman, S., Rosette, A. S., & Thompson, L. (2006). The three faces of Eve: Strategic displays of positive, negative, and neutral emotions in negotiations. *Organizational Behavior and Human Decision Processes*, 99(1), 81–101. <https://doi.org/10.1016/j.obhdp.2005.08.003>
- Kopp, C., Korb, K. B., & Mills, B. I. (2018). Information-theoretic models of deception: Modelling cooperation and diffusion in populations exposed to 'fake news'. *PLOS ONE*, 13(11), e0207383. <https://doi.org/10.1371/journal.pone.0207383>
- Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F., Kircher, T., & Robertson, E. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PLoS ONE*, 3(7), e2597. <https://doi.org/10.1371/journal.pone.0002597>
- Kwak, S. S., 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. *2013 IEEE RO-MAN*, 180–185. <https://doi.org/10.1109/ROMAN.2013.6628441>
- Lakens, D. (2014). Performing high-powered studies efficiently with sequential analyses. *European Journal of Social Psychology*, 44(7), 701–710. <https://doi.org/10.1002/ejsp.2023>
- LeBeau, B. (2019). *Power analysis by simulation using R and simglm* [Preprint]. <https://doi.org/10.17077/f7kk-6w7f>
- Lee, K. M., Jung, Y., Kim, J., & Kim, S. R. (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. *International Journal of Human Computer Studies*, 64(10), 962–973. <https://doi.org/10.1016/j.ijhcs.2006.05.002>
- Lee, M., Ahn, H. S., Kwon, S. K., & Kim, S. (2018). Cooperative and competitive contextual effects on social Cognitive and empathic neural responses. *Frontiers in Human Neuroscience*, 12, 218. <https://doi.org/10.3389/fnhum.2018.00218>
- Lench, H. C., Flores, S. A., & Bench, S. W. (2011). Discrete emotions predict changes in cognition, judgment, experience, behavior, and physiology: A meta-analysis of experimental emotion elicitations. *Psychological Bulletin*, 137(5), 834–855. <https://doi.org/10.1037/a0024244>
- Lenth, R. (2020). *emmeans: Estimated marginal means, aka least-squares means*. (R package version 1.4.7.) [Computer software]. <https://CRAN.R-project.org/package=emmeans>.
- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, 66(1), 799–823. <https://doi.org/10.1146/annurev-psych-010213-115043>
- Maggioni, M. A., & Rossignoli, D. (2021). If it looks like a human and speaks like a human ... Dialogue and cooperation in human-robot interactions. *Dialogue and Cooperation in Human-Robot Interactions*, <http://arxiv.org/abs/2104.11652>.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochemia Medica*, 22(3), 276–282. <https://hrca.srce.hr/89395>
- Mobbs, D., Weiskopf, N., Lau, H. C., Featherstone, E., Dolan, R. J., & Frith, C. D. (2006). *The kuleshov effect: The influence of contextual framing on emotional attributions*. 12.



- Moisan, F., ten Brincke, R., Murphy, R. O., & Gonzalez, C. (2018). Not all prisoner's dilemma games are equal: Incentives, social preferences, and cooperation. *Decision*, 5(4), 306–322. <https://doi.org/10.1037/dec0000079>
- Mokros, A., Menner, B., Eisenbarth, H., Alpers, G. W., Lange, K. W., & Osterheider, M. (2008). Diminished cooperativeness of psychopaths in a prisoner's dilemma game yields higher rewards. *Journal of Abnormal Psychology*, 117(2), 406–413. <https://doi.org/10.1037/0021-843X.117.2.406>
- Moors, A., Ellsworth, P. C., Scherer, K. R., & Frijda, N. H. (2013). Appraisal theories of emotion: State of the Art and Future development. *Emotion Review*, 5(2), 119–124. <https://doi.org/10.1177/1754073912468165>
- Munafò, M. R. (2016). Open Science and research reproducibility. *Eccancermedicalscience*, 10(ed56), 1–3. <https://doi.org/10.3332/ecancer.2016.ed56>
- Murphy, R. O., & Ackermann, K. A. (2014). Social Value Orientation: theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, 18(1), 13–41. <https://doi.org/10.1177/1088868313501745>
- Murphy, R. O., Ackermann, K. A., & Handgraaf, M. (2011). Measuring social value orientation. *Ssrn*, 6(8), 771–781. <https://doi.org/10.2139/ssrn.1804189>
- Novak, D., Nagle, A., Keller, U., & Riener, R. (2014). Increasing motivation in robot-aided arm rehabilitation with competitive and cooperative gameplay. *Journal of NeuroEngineering and Rehabilitation*, 11(1), 64. <https://doi.org/10.1186/1743-0003-11-64>
- Odekerken-Schröder, G., Mele, C., Russo-Spena, T., Mahr, D., & Ruggiero, A. (2020). Mitigating loneliness with companion robots in the COVID-19 pandemic and beyond: An integrative framework and research agenda. *Journal of Service Management*, 31(6), 1149–1162. <https://doi.org/10.1108/JOSM-05-2020-0148>
- Paeng, E., Wu, J., & Jr, J. C. B. (2016). Human-Robot Trust and Cooperation through a game theoretic framework. *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, 4246–4247.
- Perc, M., Jordan, J. J., Rand, D. G., Wang, Z., Boccaletti, S., & Szolnoki, A. (2017). Statistical physics of human cooperation. *Physics Reports*, 687, 1–51. <https://doi.org/10.1016/j.physrep.2017.05.004>
- Perc, M., Ozer, M., & Hojnik, J. (2019). Social and juristic challenges of artificial intelligence. *Palgrave Communications*, 5(1), 61. <https://doi.org/10.1057/s41599-019-0278-x>
- Pletzer, J. L., Balliet, D., Joireman, J., Kuhlman, D. M., Voelpel, S. C., Van Lange, P. A. M., & Back, M. (2018). Social Value Orientation, expectations, and cooperation in social dilemmas: A meta-analysis. *European Journal of Personality*, 32(1), 62–83. <https://doi.org/10.1002/per.2139>
- Pocock, S. J. (1977). Group sequential methods in the design and analysis of clinical trials. *Biometrika*, 64(2), 191–199. <https://doi.org/10.1093/biomet/64.2.191>
- Pothos, E. M., Perry, G., Corr, P. J., Matthew, M. R., & Bussemeyer, J. R. (2011). Understanding cooperation in the prisoner's dilemma game. *Personality and Individual Differences*, 51(3), 210–215. <https://doi.org/10.1016/j.paid.2010.05.002>
- Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8), 413–425. <https://doi.org/10.1016/j.tics.2013.06.003>
- Rapoport, A. (1967). A note on the “index of cooperation” for prisoner's dilemma. *Journal of Conflict Resolution*, 11(1), 100–103. <https://doi.org/10.1177/002200276701100108>
- Rapoport, A., Chammah, A. M., & Orwant, C. J. (1965). *Prisoner's dilemma: A study in conflict and cooperation* (Vol. 165). University of Michigan press.
- R Core Team. (2020). *R: a language and environment for statistical computing [Internet]* (4.0.0) [Computer software]. Foundation for Statistical Computing.
- Rick, S., & Loewenstein, G. F. (2008). The role of emotion in economic behavior. In *Handbook of Emotions* 3rd ed., 138–158. <http://www.ssrn.com/abstract=954862>
- Riek, L. D., Adams, A., & Robinson, P. (2011). Exposure to cinematic depictions of robots and attitudes towards them. *IEEE Conference on Human-Robot Interactions, Workshop on Expectations and Intuitive Human-Robot Interaction* (Vol. 6).
- Roseman, I. J., & Smith, C. A. (2001). Appraisal theory. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, methods, research* (pp. 3–19). Oxford University Press.
- Sandoval, E. B., Brandstetter, J., Obaid, M., & Bartneck, C. (2016). Reciprocity in human-robot interaction: A quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics*, 8(2), 303–317. <https://doi.org/10.1007/s12369-015-0323-x>
- Seo, S. H., Geiskkovitch, D., Nakane, M., King, C., & Young, J. E. (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 Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*. <https://doi.org/10.1145/2696454.2696471>
- Stock-Homburg, R. (2022). Survey of emotions in human–Robot Interactions: Perspectives from robotic Psychology on 20 years of research. *International Journal of Social Robotics*, 14, 389–411. <https://doi.org/10.1007/s12369-021-00778-6>
- Terada, K., & Takeuchi, C. (2017). Emotional expression in simple line drawings of a robot's face leads to higher offers in the ultimatum game. *Frontiers in Psychology*, 8, 1–9. <https://doi.org/10.3389/fpsyg.2017.00724>
- Torta, E., Van Dijk, E., Ruijten, P. A. M., & Cuijpers, R. H. (2013). The ultimatum game as measurement tool for anthropomorphism in human-robot interaction. In G. Herrmann, M. J. Pearson, A. Lenz, P. Bremner, A. Spiers, & U. Leonards (Eds.), *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (pp. 209–217). Springer. [https://doi.org/10.1007/978-3-319-02675-6\\_21](https://doi.org/10.1007/978-3-319-02675-6_21)
- Tulk, S., & Wiese, E. (2018). Social decision making with humans and robots: Trust and approachability mediate economic decision making [preprint]. *PsyArXiv*, 1–5. <https://doi.org/10.31234/osf.io/4aj8v>
- Van Dijk, E., Van Beest, I., Van Kleef, G. A., & Lelieveld, G. J. (2018, January). Communication of anger versus disappointment in bargaining and the moderating role of power. *Journal of Behavioral Decision Making*, 632–643. <https://doi.org/10.1002/bdm.2079>
- Van Dijk, E., Van Kleef, G. A., Steinel, W., & Beest, I. (2008). A social functional approach to emotions in bargaining: When communicating anger pays and when it backfires. *Journal of Personality and Social Psychology*, 94(4), 600–614. <https://doi.org/10.1037/0022-3514.94.4.600>

- Van Kleef, G. A. (2009). How emotions regulate social life. *Current Directions in Psychology*, 18(3), 184–188. <https://doi.org/10.1111/j.1467-8721.2009.01633.x>
- Van Kleef, G. A., De Dreu, C. K. W., Pietroni, D., & Manstead, A. S. R. (2006). Power and emotion in negotiation: Power moderates the Interpersonal Effects of Anger and Happiness on concession making. *European Journal of Social Psychology*, 36(4), 557–581. <https://doi.org/10.1002/ejsp.320>
- Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2004). The interpersonal effects of emotions in negotiations: A motivated Information processing approach. *Journal of Personality and Social Psychology*, 87(4), 510–528. <https://doi.org/10.1037/0022-3514.87.4.510>
- Van Kleef, G. A., De Dreu, C. K. W., & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. In *Advances in Experimental Social Psychology*, 42, 45–96. Elsevier. [https://doi.org/10.1016/S0065-2601\(10\)42002-X](https://doi.org/10.1016/S0065-2601(10)42002-X)
- Viola, T. W., Niederauer, J. P. O., Kluwe-Schiavon, B., Sanvicente-Vieira, B., & Grassi-Oliveira, R. (2019). Cocaine use disorder in females is associated with altered social decision-making: A study with the prisoner's dilemma and the ultimatum game. *BMC Psychiatry*, 19(1), 211. <https://doi.org/10.1186/s12888-019-2198-0>
- Wu, J., Paeng, E., Linder, K., Valdesolo, P., & Boerkoel, J. C. (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>
- Wykowska, A., Chaminade, T., & Cheng, G. (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1693), 20150375. <https://doi.org/10.1098/rstb.2015.0375>
- Yang, G.-Z., Nelson, J., Murphy, B., Choset, R. R., Christensen, H., H., H., Collins, S., Dario, P., Goldberg, K., Ikuta, K., Jacobstein, N., Kragic, D., Taylor, R. H., & McNutt, M. (2020). Combating COVID-19—The role of robotics in managing public health and infectious diseases. *Science Robotics*, 5(40), eabb5589. <https://doi.org/10.1126/scirobotics.abb5589>