



Universiteit
Leiden
The Netherlands

Physiological synchrony in the context of cooperation: Theoretical and methodological considerations

Behrens, F.

Citation

Behrens, F. (2020, October 28). *Physiological synchrony in the context of cooperation: Theoretical and methodological considerations*. Retrieved from <https://hdl.handle.net/1887/137983>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/137983>

Note: To cite this publication please use the final published version (if applicable).

Cover Page



Universiteit Leiden



The handle <http://hdl.handle.net/1887/137983> holds various files of this Leiden University dissertation.

Author: Behrens, F.

Title: Physiological synchrony in the context of cooperation: Theoretical and methodological considerations

Issue Date: 2020-10-28

CHAPTER 2

Under the umbrella of prosocial behavior: A critical comparison of paradigms

ABSTRACT

Despite the discontent, cruelty, and warfare that fill the daily news, people show tremendous capacities to help and cooperate with others. Prosocial behavior is used as an umbrella term capturing the diversity of selfless acts. As such, researchers have developed a variety of tasks and it is crucial to verify that they measure the same underlying construct of prosocial behavior. Previous studies have focused on comparing anonymous, one-shot economic games providing evidence for behavioral consistency across games. The current study extends these findings by (i) comparing both repeated economic and naturalistic interactive games in a within-subject design, and (ii) letting participants play in face-to-face dyadic settings. In total, 74 participants completed six tasks: three variants of a social dilemma game, an Egg Hunt game measuring helping behavior, a group decision-making paradigm requiring communication skills, and a Tangram game where participants solved puzzles together. A Principal Component Analysis revealed that two components best describe the behavior in these tasks. The three social dilemma games loaded on the first component, termed “social dilemma games”. These games were distinct from the interactive games and the helping and decision-making tasks loaded on the second component, termed “naturalistic games”. The Tangram game was unrelated to all other games. These findings suggest that the behavioral consistency observed in economic games has its limits to generalize to other types of tasks and emphasizes the importance of choosing the appropriate (combination of) paradigms to measure prosocial behavior. Theoretical and methodological differences between tasks are discussed to explain these findings.

Based on: Behrens, F., & Kret, M.E. (2020). Under the umbrella of prosocial behavior – a critical comparison of paradigms. *PsyArXiv*.

INTRODUCTION

Prosocial behavior is one of the pillars of human society. Thousands of dollars are donated every day to strangers the donor will never see, leaders from 195 countries agreed on the Paris Agreement to fight climate change, and researchers from around the world form collaborations to advance our understanding of the human mind. Prosocial behavior is used as an umbrella term incorporating a “broad range of actions intended to benefit one or more people other than oneself— behaviors such as helping, comforting, sharing, and cooperating” (Batson & Powell, 2003, p. 463). In order to understand how prosocial behavior arises, it is of pivotal importance to elucidate its underlying mechanisms, detect individual differences, and highlight situations where it is common as well as those where it is a rarity. To that extent, previous researchers have developed multiple paradigms to investigate prosocial behavior. One question that arises given the variety of tasks and the broad definition of the term “prosocial behavior” is whether behavioral tendencies are consistent across these paradigms. The few studies that have addressed this question generally observed similar behavior across tasks, reflecting a person’s overarching, temporally stable prosocial preferences (Blanco, Engelmann, & Normann, 2010; Böckler, Tusche, & Singer, 2016; Epstein, Peysakhovich, & Rand, 2016; Peysakhovich, Nowak, & Rand, 2014; Volk, Thöni, & Ruigrok, 2012; Yamagishi et al., 2013). However, these studies have mostly concentrated on anonymous, one-shot economic games, as discussed below. Although such games provide valuable insights in controlled laboratory settings, whether similar patterns emerge in more naturalistic settings remains less understood. Taking a step in this direction, the current study aims to extend the previous findings by considering a more heterogeneous set of paradigms including both economic games and three more naturalistic tasks, and by playing iterated games in face-to-face dyadic interaction settings.

In the fields of economics, psychology, political sciences and biology it is common to measure prosocial behavior by means of social dilemma games. A strength is that the social dilemmas are, for the neurotypical population, easy to comprehend, yet tap into complex motives, cognition and emotions. These games are designed to reflect real-life decision-making problems with conflicting self- or collective interests. Specifically, such games have been developed to measure different subtypes of prosocial behavior, including but not limited to cooperation (e.g., Prisoner’s Dilemma game and Public Good game), trust (e.g., Trust game), generosity (e.g., Dictator game). Although they are designed to measure different aspects of prosocial behavior, behavior is believed to be driven by a person’s general prosocial preference or “phenotype” (Balliet, Parks, & Joireman, 2009; Peysakhovich et al., 2014; Poncela-Casasnovas et al., 2016). In line with this, previous studies have shown that participants behave quite consistently across a variety of tasks. For example, Yamagishi et al. (2013) compared five economic games (two variants of the Prisoner’s Dilemma game, Trust game, Dictator game, and a Faith game) and showed that behavior across these games were correlated (except for the Dictator game). Similarly, Peysakhovich and colleagues (2014) observed that participants who cooperated in the Public Good game were also more likely to allocate more resources to their partner during the Trust game and the Dictator game. Importantly, although these cooperation games showed moderate correlations, suggesting they tap into a similar construct, they were distinct from norm-based games such as the Ultima-

tum game and the Punishment game. In other words, participants who cooperated in one game, were more likely to cooperate in another, but were not necessarily inclined to punish others for non-cooperative decisions. This distinction was further supported by Böckler et al. (2016), who compared 14 computerized tasks including economic games, hypothetical distribution tasks, and self-reported measures. Applying a Principal Component Analysis, the authors showed that games measuring altruistic-motivated prosocial behavior, such as donating and helping, clustered together; however, were distinct from norm-based punishment games and self-reported prosocial behavior. In sum, these studies support the notion that social dilemma games measuring different aspects of prosocial behavior generally tap into similar overarching behavioral tendencies.

The question remains whether such clustered behavior is also evident when using more heterogeneous paradigms. Böckler and her colleagues (2016) have made an attempt in this direction by including a broader range of paradigms other than economic games. However, these games were still based on straightforward pay-off structures where people indicate their decisions by pressing keys in an isolated room. Given that prosocial behavior is social by nature, it is crucial to investigate whether the above described findings also apply to situations where people interact with each other in a face-to-face setting and play games with an uncertain pay-off structure.

To that end, in the current study we compared behavior in six prosocial behavior games: three variants of a social dilemma game and three more naturalistic games, where people were asked to build puzzles together (Vink, Hasselman, Cillessen, Wijnants, & Bosman, 2018), communicate and exchange information (Nevicka, Ten Velden, de Hoogh, & van Vianen, 2011), and help each other in collecting eggs (McClung, Placi, Bangerter, Clément, & Bshary, 2017). With regard to the social dilemma games, we tested three variants of the Prisoner's Dilemma game: (1) the classical variant with two response options (cooperate or defect); (2) an extended version of the Prisoner's Dilemma game, where the pay-off structure was extended from a 2×2 to a 6×6 matrix; and (3) an adjusted version developed to test children and chimpanzees, where participants can decide to pull a rope (i.e., cooperate) or not (i.e., defect) (Sánchez-Amaro, Duguid, Call, & Tomasello, 2019). Including these three variants allowed us to investigate whether participants based their decision on the same principles across games despite changes in the scale (six versus two response options) and the way of indicating a decision (key press versus rope pull).

The three additional games are less commonly used paradigms to measure prosocial behavior, yet tap into related processes. First, the Tangram game requires participants to build puzzles together (Vink et al., 2018). The more efficiently participants work together, the more puzzles they can complete. In contrast to other games described in this study, this task required people to physically work together on one problem such that participants could complement their actions to complete the puzzles. The second paradigm was the Hidden Profile game which was originally designed to study group decision-making processes rather than prosocial behavior (Nevicka et al., 2011; Stasser & Titus, 1985). In this game, participants are asked to find the most suitable job candidate, by sharing information. Crucially, performance during this game depends on individuals' skills to cooperate and share partly *unique* information (Wittenbaum, Hollingshead, & Botero, 2004). If participants only consider shared information, they are led to an incorrect decision. However, when all unshared information is exchanged, the correct "hidden" decision is revealed. Thus, this task measures how well people communicate with each other and reach the

goal of finding the best job candidate. The third game was the Egg Hunt game measuring helping behavior (McClung et al., 2017). Specifically, the behavior of interest was how much effort participants put into helping others collect their eggs at the cost of the limited time to collect their own. This task still loosely incorporates the idea of the social dilemma because a person can choose between maximizing his or her own rewards at the expense of the joint outcome. However, this structure was neither directly visible nor highlighted in the instructions and helping behavior could develop in a natural environment.

By incorporating a variety of tasks ranging from different versions of the classical controlled social dilemma paradigms, to interaction tasks reflecting more natural and less abstract situations, we aimed to answer the questions: How robust is the consistent behavioral tendency observed across different games in previous studies? Are some games more related to each other than other games? We expected similar behavior among the three variants of the social dilemma game, because they are based on the same principles and similar pay-off structures. With regard to the three naturalistic games, the expectations were less straightforward as to how they would correlate with each other and with the social dilemma games. The social dilemma games incorporate clear response options and participants could choose to act prosocially or not. In contrast, in the three naturalistic games the options were less clear-cut and the *willingness* to act prosocially was constrained by the *ability* to do so. For example, the classical Prisoner's Dilemma game requires a certain amount of cognitive abilities to understand the abstract payoff structure. The Tangram game depends on spatial skills and the Hidden Profile game addresses communication abilities. Differences in these abilities could undermine the correlation in behavior between the games. On the other hand, as described above, prosocial preferences have been demonstrated to show across a range of tasks and subtypes of prosocial behavior have been shown to relate to similar underlying motivational tendencies. For example, prosocial people as classified by their social value orientation have been shown to cooperate more in the social dilemma game and show more helping behavior than proself individuals (C. Boone, Declerck, & Kiyonari, 2010; Van Lange, Bekkers, Schuyt, & Van Vugt, 2007). Similarly, social value orientation has also been related to motivational processes in the Hidden Profile game (De Dreu, Nijstad, & Van Knippenberg, 2008). Furthermore, behavior in economic games have been shown to translate to behavior outside the lab, suggesting that behavior should also translate to the more naturalistic tasks used in the current study (Benz & Meier, 2008; Böckler et al., 2016; Franzen & Pointner, 2013). These arguments suggest that participants would show similar behavioral tendencies across all six games.

METHOD

Participants

In total, 74 individuals participated in this study, completing six separate experimental tasks ($M_{age} = 22.05$, $SD_{age} = 2.55$, $Range_{age} = 18-31$ years). The subject-to-items ratio (13:1) was good which has been shown to be important when performing a Principal Component Analysis (Osborne & Costello, 2004). Participants were recruited via an online recruitment system, flyers in the University building, or personal contacts. More demographical information about the participants

are listed in Appendix A1 (see Table A.S1). Participants received written and oral instructions in Dutch or English. All participants provided written informed consent after receiving information about the study and prior to the start of the experiment. Due to participants arriving late or leaving early during the experiment and due to technical issues, the number of participants per game differed between 53 and 73. The study was approved by the Psychology Research Ethics Committee of Leiden University (CEP19-0318/223).

Design

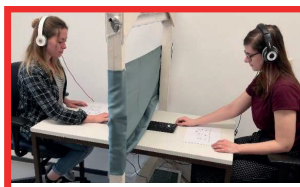
The study was a within-subject design with all participants playing each of the six games (with the exception of those arriving late/leaving early). All games were played in dyads. To minimize the influence of the partner on a person's prosocial behavior, the dyadic composition changed after two games. The two games that a dyad played together as well as the order of the games was counterbalanced between groups. The only restriction was that the Prisoner's Dilemma game was played as one of the first two games and the extended Prisoner's Dilemma game as one of the last two games. This restriction was introduced for two reasons: (1) the extended Prisoner's Dilemma game builds on the original Prisoner's Dilemma game, making it easier to instruct participants on the more complex version. (2) Compared to the other tasks, the instructions were more complex, making these two games cognitively more demanding. Therefore, having a break in between would increase participants' level of concentration.

Material

All participants played six games with three different partners (Figure 1). They were not allowed to talk about the (strategy of the) games, but could chat about task-unrelated topics (monitored by the experimenter who was either involved in administering the task or stood close to the participants). This restriction was not applied during the Hidden Profile game, as the task was to have a discussion.

Prisoner's Dilemma game. In this two-choices game, participants can choose to cooperate or defect with their partner and the amount of points a person receives depends on the own and the partner's choice. When both players cooperate, the outcome is higher compared to when both defect and gives the highest joint outcome. However, independent of what the partner chooses, the higher individual outcome is always achieved by choosing to defect (Table 1). A player receives the lowest amount of points when s/he cooperates and the other player defects. Thus, a dilemma is created between maximizing the individual or the joint outcome. In the current study, participants played three practice and ten experimental trials. Practice trials were included to ensure that participants understood and were familiarized with the game. These trials were not included in the analysis. Auditory pre-recorded instructions were provided via headphones (for a similar approach, see Behrens & Kret, 2019). The sequence of a trial was as follows: Participants were asked to look at each other (participants could only see each other's face). After four seconds, participants were asked to look down and indicate their decision (i.e., cooperate or defect) on a keyboard. The response window was three seconds. Afterwards, they heard that they

both had made a choice and the next trial began. Participants did not receive any feedback about their decisions during the game. The instructions were phrased to choose between option A and option B, avoiding any references to cooperation and defection. The face-to-face interval was introduced to allow nonverbal communication which helps participants to read the intentions of their partner (Behrens & Kret, 2019). The measure of interest was the proportion of decisions to cooperate (i.e., number of decisions to cooperate divided by the total number of trials). Two participants were excluded because they had more than four out of ten missing values.



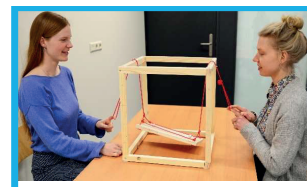
Prisoner's Dilemma

- Choose to cooperate or defect
- Before each decision, participants look at each other
- Dichotomous scale (1 = cooperate)
- 10 trials
- Duration: 5 min
- Mol: % cooperative decisions

		The number of points to which you get					
		A	B	C	D	E	F
You	A	4,0	3,6	3,6	3,6	3,2	3,0
	B	3,6	5,4	3,2	3,0	2,8	2,6
	C	3,2	3,0	2,8	2,6	2,4	2,2
	D	2,8	2,6	2,4	2,2	2,0	1,8
	E	2,4	2,2	2,0	1,8	1,6	1,4
	F	2,0	1,8	1,6	1,4	1,2	1,0
		A	B	C	D	E	F

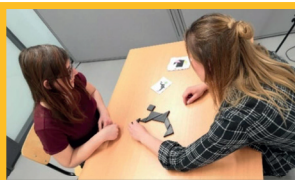
Extended Prisoner's Dilemma

- Extends Prisoner's Dilemma to a 6 x 6 payoff structure
- 6-point scale (6 = fully cooperate)
- 10 trials
- Duration: 5 min
- Mol: mean of points (between 1 – 6) / 6



Rope Pull

- Based on the principles of the Prisoner's Dilemma
- Pull (=cooperate) or not (=defect)
- Dichotomous scale (1 = cooperate)
- 10 trials
- Duration: 3 min
- Mol: % cooperative decisions



Tangram Puzzle

- Complete as many puzzles as possible
- Two conditions: complete puzzles cooperatively and alone
- Duration: 2 x 5 min
- Mol: completed puzzles (coop) / completed puzzles (total)



Hidden Profile

- Find best job candidate with different characteristics
- Un-/Shared information in profiles
- Combined information reveals best candidate
- Duration: 7 min (5 min discussion)
- Mol: % unique information exchanged



Egg Hunt

- Participants are rewarded for different colors of eggs
- Collecting eggs is time-costly
- Do participants help each other?
- Duration: 7 min (5 min egg hunting)
- Mol: % helping behavior when finding egg from other participant

Figure 1. Information about the set-up of the six prosocial behavior games played by each participant; MoI = Measure of Interest included in the analysis.

Table 1*The payoff matrix of the Prisoner's Dilemma game*

		Other	
You		C	D
C		3-3	1-4
D		4-1	2-2

Note. The first number refers to the points earned by "You".

Extended Prisoner's Dilemma game. This game extends the original Prisoner's Dilemma game from a 2×2 to a 6×6 payoff structure and was adopted from Behrens et al. (2019). Thus, the response options changed from a dichotomous choice (i.e., cooperate or defect) to a six-point scale ranging from option A (=fully defect) to option F (=fully cooperate). Apart from this adjustment in the payoff structure, the procedure was the same as for the classical Prisoner's Dilemma game described above. The measure of interest was the mean willingness to cooperate on the 6-point scale (1=fully defect [option A] to 6=fully cooperate [option F]). To make the measure comparable to the other games, we transformed the mean value to a proportion value.

Rope Pull game. This game was adopted from Sánchez-Amaro and colleagues (2019) who used this game to measure cooperation in chimpanzees and children. The game is based on the principles of the Prisoner's Dilemma game, but instead of pressing keys, participants could pull a rope or not. When both participants pulled the rope (i.e., mutual cooperation), a tray with two rewards was lifted and both participants received one reward each. If only one participant pulled (i.e., unilateral cooperation), the tray was lifted on that participant's side and the two rewards rolled to the side of the other participant. If nobody pulled for 15 seconds (i.e., mutual defection), no one received a reward. Hence, the principle is similar to the Prisoner's Dilemma game in that mutual cooperation is more beneficial for both participants than mutual defection and that with unilateral cooperation the defector earns more points than through mutual cooperation. However, the game differs such that mutual defection and unilateral cooperation result in the same outcome (no reward). Compared to the payoff structure of the current Prisoner's Dilemma game another difference was that mutual cooperation led to the same reward than a turn-taking strategy (unilateral cooperation by one participant and then by the other; Sánchez-Amaro et al., 2019). Participants first read instructions about the game and subsequently the experimenter showed the four possible outcomes to the participants with the apparatus. Afterwards, participants played ten trials and the experimenter recorded the outcome of each trial on a sheet of paper. As was the case for all other tasks as well, the game was phrased in a neutral way without referring to cooperation and defection. The measure of interest was the proportion of times a participant cooperated out of the total number of trials. The willingness to cooperate was based on the reward distribution such that a person cooperated if both participants received one reward each, or when the other participant received both rewards.

Tangram game. In this game, participants are asked to make predefined figures with a number of puzzle pieces that have different shapes (e.g. triangle, square; Vink et al., 2018). In the current study, the same seven pieces were used to make different figures. Participants were asked to complete as many figures as possible within five minutes. They were only allowed to continue with the next figure after completing the previous one which was checked by the experimenter. Participants played two rounds, once individually and once together. The individual condition served as a baseline measure of their skill in performing the task. The measure of interest was the difference in completed puzzles between the cooperative and individual condition (completed puzzles in the cooperative condition minus completed puzzles in the individual condition). Additionally, we explored different measures of prosocial behavior on the dyadic level which are presented in Appendix A3. We made three sets of figures with eighteen figures each that were individually printed in black with only the outline of the figure visible. Which set was given in the individual and cooperative condition was counterbalanced. The order of the figures within each set was kept constant. The difficulty of the figures was based on the performance of the experimenters prior to data collection. For some figures, the outline of one or two puzzle pieces were shown in a different color to make it easier.

Hidden Profile game. This game was adopted from Nevicka and colleagues (2011) and was originally designed to measure group decision-making processes. Here, participants need to find the most suitable candidate for a job. Each participant receives a profile with information about the three potential candidates. What participants do not know is that some information is shared between them, whereas other information is unique for each participant. Based on the (incomplete) information of each profile, a suboptimal candidate stands out; however, after combining all information from the different profiles, another candidate is more suited (has the most positive and least negative characteristics). Thus, the true best candidate is “hidden” in the unshared information distributed among the profiles. In the current study, we used the profiles from Nevicka and colleagues (2011) where participants needed to find the best candidate for a secret agent position with validated positive, neutral and negative characteristics. To adjust the original three-player game to the dyadic setting of the current study, we excluded one characteristic per candidate. Both profiles (one for each participant) consisted of nine characteristics for each candidate, of which six were shared among the profiles. Eight additional characteristics per candidate were evenly distributed between the two profiles, that is, each profile included four unique characteristics per candidate. In the shared information, candidate A had three neutral and three negative characteristics, candidate B had six positive characteristics and candidate C had three positive and three negative characteristics. Based on this information, candidate B would be the preferred candidate. In the unshared information, candidate A had eight positive characteristics, candidate B had two neutral and six negative characteristics, and candidate C had two positive and six neutral characteristics. Therefore, after combining the information candidate A was most suited. Examples of the characteristics were: “can read code language” (positive), “is 180 cm tall” (neutral), and “is afraid of heights” (negative). During the game, participants first read the profiles and made an individual decision about which candidate they would choose.

Subsequently, participants had a five-minute discussion and then made a joint decision about the best candidate. The performance of a dyad was operationalized by the number of unshared characteristics that were exchanged during the discussion. To measure this, participants filled out a checklist with the characteristics of the candidates and indicated whether each of the items described the candidates. A characteristic was considered “exchanged” if the participant who did not have that characteristic in his/her profile indicated that it belonged to the corresponding candidate. The measure of interest was the proportion of exchanged characteristics in relation to the total number of unshared information of a participant’s profile.

Egg Hunt game. The Egg Hunt game was adopted from McClung and colleagues (2017) who used this game to measure helping behavior. The idea is that participants are assigned different colors of eggs and earn rewards for each egg of their color they collect during the egg hunt. However, there is not enough time to collect all own eggs and collecting an egg is time-costly. Thus, the behavior of interest is what participants decide to do when they find an egg of the partner’s color: Does the participant invest the time to help the partner by collecting the eggs or not? In the current study, 90 eggs were wrapped in paper sandwich bags: 18 orange, 18 pink, and 54 green eggs. One participant was rewarded for all collected pink eggs, the other participant for all orange eggs. The green eggs were not rewarded. The wrapped eggs were placed in a room and participants could simultaneously hunt for the eggs. For each egg, participants first had to unwrap the paper bag, take out the egg to see the color and then (i) put the egg back in the bag, (ii) put the egg next to the bag, or (iii) take the egg, run around a chair twice (there were two chairs in the room) and then put the egg into one of two baskets (one basket for the orange eggs and another basket for the pink eggs). Participants were rewarded for the eggs in their basket with chocolate (see below). The game lasted five minutes. During the game, both participants wore eye-tracking glasses (Tobii Pro Glasses 2) recording their behavior. The behavior was coded afterwards following McClung and colleagues’ (2017) scoring scheme. Specifically, when an egg of the other participant was found, the behavior was classified into three categories: (1) costly helping: the egg is collected or passed on to the other player; (2) no costly helping: the egg is left visible to the other participant; (3) neglect: the egg is put back in the paper bag. We also added a fourth category representing the “competitive” behavior in other games: (4) active hiding: the paper bag with the egg is actively hidden. The latter behavior was, however, not evident in any video recordings. The measure of interest was the proportion of helping behavior (costly and no costly helping) from the times that an egg with the color assigned to the other participant was found. Unfortunately, some behavior could not be classified because either the participant did not look down to see the egg or the color could not be clearly identified from the video. Participants with ten or more unidentified behaviors were excluded from the analysis. Data of 53 participants could be included in the analysis.

Questionnaire. At the end of the experiment, participants filled out a questionnaire where they were asked to indicate their gender, age, nationality, highest completed education level, and the number of siblings (for descriptive statistics, see Table A.S1). Furthermore, participants indicated how much they knew and liked the three different partners on a scale from 0 – 100. Descriptive statistics of these scales are presented in Appendix A2 (see Table A.S2 and Figure

A.S1). We also present the Spearman's rho correlation between the liking rating and prosocial behavior in each game (see Table A.S3). Finally, participants were asked about what they thought the purpose of the study was.

Procedure

A group of four same-sex participants was invited into the lab. Upon arrival, participants were separated, read the information about the study and gave informed consent. They were also given a colored wristband which was used to form the dyads. Which participant received which color was based on the time of arrival (the first individual had the green wristband, the second the orange, the third the purple and the fourth the blue). When there were no more questions and all participants were ready, the first two games started. All games were played in a big lab with a separation wall in the middle so that the two dyads could play games simultaneously without seeing each other. The first two dyads were formed with participants wearing the green and orange (Dyad 1) and the purple and blue (Dyad 2) wristbands. One of the dyads started with the Prisoner's Dilemma game, the other dyad started with another game. Which dyad began with the Prisoner's Dilemma game was counterbalanced, and the second game was changed between groups. After finishing the games, the dyads switched places and played the game the other dyad just played. Then, the formation of the dyads changed (Dyad 3: green + purple; Dyad 4: orange + blue) and the next round of two games were played. In the last round, the last two dyads were formed (Dyad 5: green + blue; Dyad 6: orange + purple) and the extended Prisoner's Dilemma game and the last remaining game were played. Thus, all participants played six games with three different partners. Because the order of the games was changed between the groups and therefore the combination of games that were played within dyads, order effects were negligible and the dependency between observations on a dyadic level was minimalized (e.g., five out of 19 groups played the Prisoner's Dilemma game paired with the Egg Hunt game in one round). After that, participants filled out the questionnaire (see Materials). Finally, participants received as many M&Ms as they earned points during the Egg Hunt, Rope Pull, and Tangram games (Egg Hunt: the number of eggs collected of the participant's color; Rope Pull: number of rewards received; Tangram: number of completed puzzles in both conditions). Depending on their preference, participants were additionally paid with course credits or money and thanked for participating.

Statistical Analysis

For each game, one measure of interest was calculated per participant as described in the Material section. In the first step, descriptive statistics of each game are presented. Second, correlation coefficients and 95%-confidence intervals for each combination of games are reported. Given that most measures were not normally distributed, we report Spearman's rho correlations. In the third step, we conducted a Principal Component Analysis (PCA) to investigate which tasks could be described as measuring similar behavioral tendencies. The following settings were used: To determine the number of components, the parallel analysis was applied where a component was selected when the eigenvalue of that component was larger than the parallel average random

eigenvalue. For the rotation method, the oblique rotation method *promax* was chosen, because this method allows the components to be correlated. In an additional step, we further explored differences between the social dilemma games by looking at the absolute cooperation rates both on the individual level (i.e., a person's willingness to cooperate) and on the dyadic level (i.e., mutual cooperation). These results are reported in Appendix A4. A significance level of $\alpha = 0.05$ was used and analyses were performed in JASP 0.10 (JASP-Team, 2019) and SPSS (Version 25).

RESULTS

Descriptive statistics

Descriptive statistics of each prosocial behavior measure are presented in Table 2 and displayed in Figure 2. Except for the Tangram game, the prosocial behavior is operationalized as proportions. For the Tangram game, the difference in completed figures between the cooperative and individual conditions was calculated. A positive value therefore corresponds to more completed figures in the cooperative condition.

The highest cooperation rate was observed in the Rope Pull game with almost .85. The minimum was 0.4 meaning that participants cooperated at least four out of ten times. This was not the case in the two Prisoner's Dilemma games where some people also defected at all times. The cooperation rates of the two Prisoner's Dilemma games and the Egg Hunt game were in a similar range (around .60). The lowest rate was observed in the Hidden Profile game with only .33.

Table 2

Descriptive statistics of the proportion of prosocial behavior for the six games

Game	Mean	Median	SD	MAD	Maximum	Minimum	Range	N	Missing
Prisoner's Dilemma	.61	.60	.33	.30	1.00	.00	1.00	70	4
Extended Prisoner's Dilemma	.69	.68	.26	.23	1.00	.17	.83	70	4
Rope Pull	.84	.90	.19	.10	1.00	.40	.60	72	2
Tangram	0.10	0.00	1.80	1.00	5.00	-5.00	10.00	73	1
Hidden Profile	.33	.33	.29	.25	1.00	.00	1.00	70	4
Egg Hunt	.60	.86	.45	.14	1.00	.00	1.00	53	21

Note. SD = standard deviation; MAD = median absolute deviation; N = sample size.

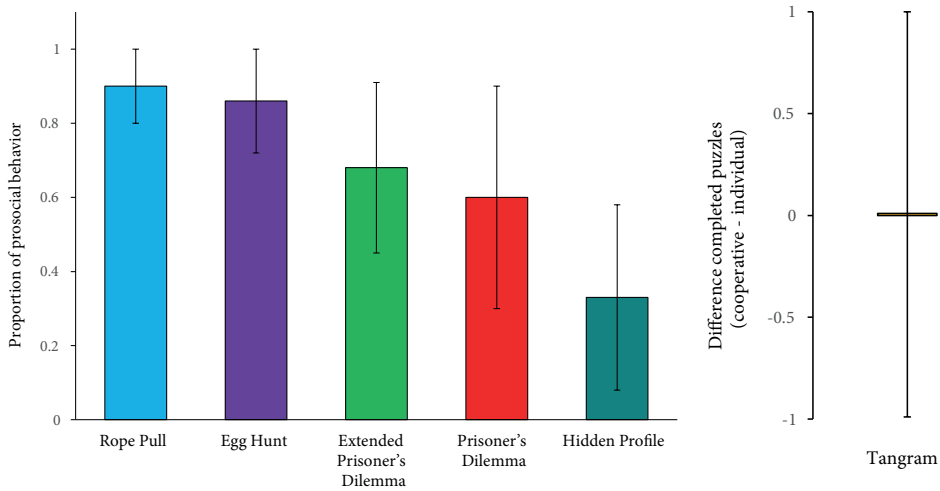


Figure 2. Median (\pm Median Absolute Deviation) of the proportion of prosocial behavior per game (left). The Tangram game (right) is displayed separately because it is based on a scale between -5 and 5, whereas all other measures are based on proportions.

Correlations and Principal Component Analysis

The bivariate correlation coefficients of each combination of games are displayed in Table 3 and Figure 3. The strongest correlation was evident between the Prisoner's Dilemma game and its extended version. Both versions were also positively correlated with the outcome of the Rope Pull game. Thus, if a person was willing to cooperate in the Prisoner's Dilemma game, s/he was also likely to cooperate in the extended version of the Prisoner's Dilemma game and the Rope Pull game. The three games were not significantly correlated with the more naturalistic games. The correlations between the latter games were non-significant.

The Principal Component Analysis confirmed the pattern seen in the correlation matrix and revealed that the six games could be best represented by two components. The component loadings are depicted in Table 4. The Prisoner's Dilemma game, its extended version, and the Rope Pull game loaded positively on the first component. As the three games are all variants of a social dilemma game, we referred to this component as the "social dilemma games" component. The Egg Hunt game loaded positively and the Hidden Profile game loaded negatively on the second component. To reflect the distinction between the economic games of the first component and the more naturalistic games of the second component, we called the latter component "naturalistic games". The two components were slightly correlated (.03).

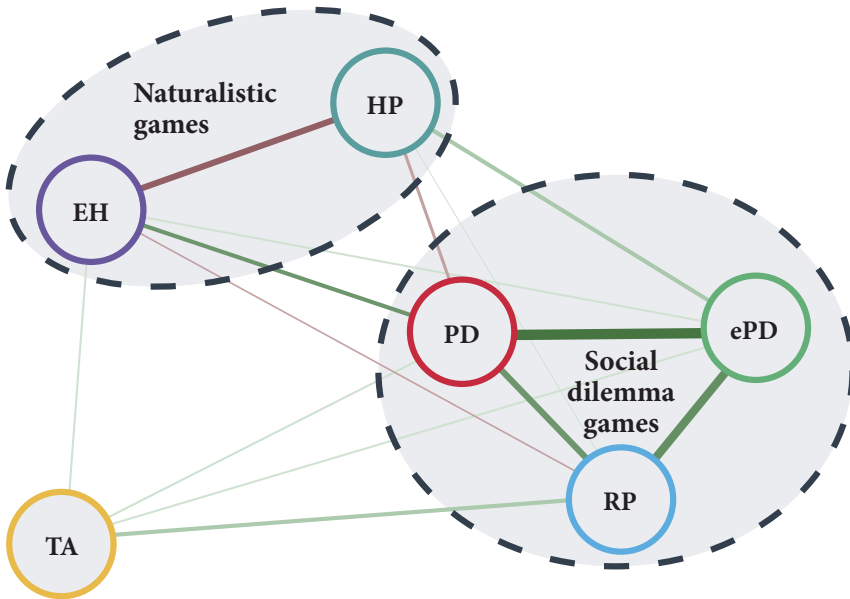


Figure 3. Visualization of the correlations between the six cooperation tasks. The thickness of the lines represents the strength of the correlation. Green lines reflect positive correlations, red lines negative ones. EG = Egg Hunt; HP = Hidden Profile; PD = Prisoner's Dilemma; ePD = extended Prisoner's Dilemma; TA = Tangram; RP = Rope Pull.

Table 3

Correlation matrix of the six games with the Spearman's rho coefficients, the corresponding 95% confidence interval and the sample size of each pair

		Prisoner's Dilemma	Extended Prisoner's Dilemma	Rope Pull	Tangram	Hidden Profile	Egg Hunt
Prisoner's Dilemma	r_s	1	.44***	.31*	.08	-.15	.22
	95%-CI		.22-.62	.08-.51	-.16-.31	-.38-.09	-.06-.47
	<i>p</i> -value		<.001	.010	.517	.222	.117
	<i>N</i>	70	66	68	69	67	52
Extended Prisoner's Dilemma	r_s		1	.34**	.05	.09	.06
	95%-CI			.11-.53	-.19-.28	-.16-.32	-.22-.33
	<i>p</i> -value			.004	.709	.487	.676
	<i>N</i>		70	69	70	67	50
Rope Pull	r_s			1	.13	.02	-.05
	95%-CI				-.11-.35	-.22-.25	-.32-.23
	<i>p</i> -value				.287	.881	.712
	<i>N</i>			72	71	69	51
Tangram	r_s				1	-.00	.04
	95%-CI					-.23-.23	-.24-.31
	<i>p</i> -value						.785
	<i>N</i>				73	70	52
Hidden Profile	r_s					1	-.24
	95%-CI						-.49-.04
	<i>p</i> -value						.088
	<i>N</i>					70	51
Egg Hunt	r_s						1
	95%-CI						
	<i>p</i> -value						
	<i>N</i>						53

Note. r_s = Spearman's rho correlation coefficient; CI = confidence interval; *N* = sample size;

* $p < .05$; ** $p < .01$; *** $p < .001$.

Table 4
Component Loadings of the Principal Component Analysis

	PC 1	PC 2	Uniqueness
Extended Prisoner's Dilemma	0.80	.	0.36
Egg Hunt	.	0.77	0.40
Hidden Profile	.	-0.77	0.41
Prisoner's Dilemma	0.72	.	0.34
Rope Pull	0.69	.	0.48
Tangram	.	.	0.96

Note. Applied rotation method is promax.

DISCUSSION

Prosociality is central to humanity's unique capacity for large-scale cooperation. Experimental paradigms that measure prosocial behavior can help us understand how it emerges, as it allows researchers to investigate its contextual boundaries, zoom in on individual differences, and take factors such as previous experiences and costs into account. Researchers have designed multiple paradigms that tap into prosocial behavior. A crucial question is whether such paradigms measure a similar construct, that is, a person's general tendency to act prosocially, or whether the different paradigms tap into distinct subcomponents of prosociality. Previous studies have shown that behavioral tendencies are fairly consistent across economic games and translate to prosocial behavior outside the lab (e.g., voluntarily filling out a feedback form, Peysakhovich et al., 2014); sending back a "misdirected letter" enclosing money, (Franzen & Pointner, 2013; Stoop, 2014; Blanco et al., 2010; Böckler et al., 2016; Epstein et al., 2016; Yamagishi et al., 2013). The current study extends these findings by investigating both economic games and more naturalistic games. Contrary to the previous studies, in all six games, participants directly interacted with their game partners, better simulating real-life interactions. Investigating whether previous findings in anonymous, one-shot economic games also translate to more ecologically valid settings scrutinizes the robustness of the previously observed consistent behavior across paradigms. Given that participants engaged in face-to-face interactions in all games, the effect of such interaction on prosocial behavior was assumed to be constant across games. Consistent with previous findings, we observed that behavior in the three variants of a social dilemma game positively correlated. However, such consistency did not generalize to the three naturalistic games as evident by negligible correlations between the economic and naturalistic games. This pattern was also apparent in a Principal Component Analysis (PCA) showing that the series of tasks split into two components, which we dub "social dilemma games" and "naturalistic games". The three variants of the social dilemma game loaded positively onto the first component, whereas the Egg Hunt game loaded positively and the Hidden Profile game loaded negatively onto the latter component. The Tangram game was not related to any of the other games. In the section below, we will discuss these results in terms of theoretical and methodological considerations.

The “social dilemma games” component of the PCA showed that behavior in the three different variations of the social dilemma game was positively related (the classical and extended Prisoner’s Dilemma game and the Rope Pull game). If participants cooperated in one of these games, they were more likely to do so in the other two as well. This was expected as all three games were designed to measure cooperative behavior. Specifically, they use similar outcome structures incorporating the essential ingredients to induce the dilemma between self- and collective interests: mutual cooperation is more beneficial for the joint outcome than mutual defection, and unilateral cooperation is more beneficial for a defector than mutual defection. The games differed with respect to the response scale (dichotomous versus 6-point scale) and the way of making a decision (pulling a rope versus pressing a button). The clustering of the three variants of the social dilemma game suggest that they are robust against these differences and capture people’s general tendency to cooperate.

The behavior shown in the social dilemma games was distinct from the behavior displayed in the three naturalistic games. In other words, the two types of tasks did not measure the same underlying prosocial preferences. This discrepancy could result from theoretical differences, suggesting that the tasks measure distinct subcomponents of prosocial behavior. It could also be that methodological issues contributed to these findings as tasks differed regarding the level of feedback and clarity of the response options and their consequences. In the following, we will discuss these possible explanations in more detail.

The distinction between the social dilemma games and the more naturalistic games might be explained from a theoretical point of view such that they capture distinct subcomponents of prosocial behavior. For example, while social dilemma games are designed to measure cooperative behavior, the Egg Hunt game measures helping behavior. The lack of a relationship in behavior between these tasks might be the result of conceptual differences: While helping behavior is one-directional with one person helping to attain the goal of another person, cooperative behavior is bidirectional and implies interdependence, that is, the success of cooperation depends on two or more people working together towards a common goal (Penner, Dovidio, Piliavin, & Schroeder, 2005). Thus, while the conflict between self- and collective interests is inherent in cooperation, it is less salient in one-directional helping. A person might therefore be willing to cooperate with another person to achieve a common goal, but not necessarily help to achieve another person’s goal. However, although conceptual differences are evident, such distinction has not been shown to elicit distinct behavior. For example, Böckler and colleagues (2016) showed that people who act prosocially during economic games also showed more helping behavior in the Zurich Prosocial Game (ZPG, Leiberg, Klimecki, & Singer, 2011). Furthermore, behavior in other games where the outcome depended on the other player (i.e., the Public Good and Trust Game) have been shown to correlate with behavior in the Dictator Game where the outcome of the dictator is independent of the other player (Peysakhovich et al., 2014; Yamagishi et al., 2013). As emphasized by Peysakhovich and colleagues (2014), these findings do not imply though that behavior across these games are driven by the same underlying motivation. For example, some games might be driven by reciprocity, while others might be driven by equality and altruistic preferences. Nevertheless, these studies suggest that despite differences in the interdependence of individuals that are inherent to the games, this likely does not explain the distinct behavior

observed in the current study. Thus, the question remains why we did not observe consistent behavior across tasks. We now turn to possible methodological considerations that might shed more light onto this question.

The aim of the current study was to extend previous studies by incorporating more naturalistic games that reflect a range of situations also encountered in real-life interactions. Stepping away from the controlled context of economic games increases the ecological validity of the tasks, but simultaneously introduces additional factors that might influence the behavior of interest. Methodological differences between tasks such as the level of feedback and transparency of the response options and their consequences might have therefore contributed to the results observed in the current study.

First of all, the level of feedback differed such that participants received no feedback about each other's decisions during the classical and extended Prisoner's Dilemma games, whereas in all other games, individuals knew how prosocially their partner acted. Feedback has been shown to increase cooperation as it provides valuable information about a person's future decisions (Behrens & Kret, 2019; Jorgenson & Papciak, 1981; Monterosso et al., 2002). In line with this, we observed that participants cooperated more during the Rope Pull game, where participants received immediate feedback about the other person's move, compared to the classical and extended Prisoner's Dilemma game, where no feedback was provided (see Appendix A4). However, despite the overall increase in cooperation rates, the relationship between the games still showed that people behaved consistently between the games. In other words, a prosocial person still acted more prosocially than a less prosocial person despite the additive effect of feedback. Thus, although feedback influenced the overall level of prosocial behavior, our results suggest that this factor did not substantially affect the relationship between the games.

Second, a crucial difference between games was how obvious the response options and their consequences were for each game. In the social dilemma games, the response options were described in a pay-off structure and a participant knew that she could either cooperate or defect depending on her own preferences. On the contrary, in the Egg Hunt game participants were not explicitly informed that they could help each other and in the Hidden Profile game it was not mentioned that some, but not all of the information that both partners received was identical (see also McClung et al., 2017). Therefore, in the social dilemma games people could make a weighted, informed choice about whether they wanted to cooperate or not; whereas in the Egg Hunt and Hidden Profile game, people needed to discover the possibility of helping and sharing unique information first before they could work together. In other words, the degree of prosocial behavior displayed in these two games might have been undermined by whether people discovered *how* they could behave prosocially. However, as described above in the case of the effect of feedback, such difference in explicitly informing about the response options could affect the overall level of prosocial behavior without affecting the relationship between games. Although it would fit the pattern that behavior in the Egg Hunt and Hidden Profile game loaded on the same component, but were distinct from the other tasks, future studies are needed to investigate this explanation more directly.

Although behavior was related in the Hidden Profile and Egg Hunt game, we were surprised to observe that the Egg Hunt loaded positively and the Hidden Profile game negatively on the “naturalistic games” components which was consistent with a marginally significant negative correlation between the two games: a person who was more helpful in the Egg Hunt game shared less information in the Hidden Profile game. One possible explanation might be attributed to differences in the underlying motives causing the behavior. While the motivation for engaging in helping behavior in the Egg Hunt game is likely to be prosocially-driven, this is not necessarily the case for information exchange in the Hidden Profile game. Although it is assumed that working together on a common goal stimulates information exchange (De Dreu et al., 2008; Toma & Butera, 2009), research has shown that people generally stick with their own *a priori* decision (Wittenbaum et al., 2004). This bias not only motivates people to share as many characteristics of their own profile to convince others of their preferred candidate, but also influences individuals’ encoding and retrieval of information from others’ that is inconsistent with their own preference (De Dreu et al., 2008). The bias can be driven by competitive motives, where people want others to adopt their opinion, or by cooperative motives, where they are genuinely convinced that the own preferred candidate serves the group interest best (Wittenbaum et al., 2004). A previous study observed more information exchange in a cooperative compared to a competitive condition suggesting that information sharing is driven by cooperative motives (Toma & Butera, 2009). However, this study investigated the relationship between cooperation and information sharing on the contextual rather than individual difference level. Specifically, they manipulated the group’s goal by emphasizing that the individuals should come to a joint decision (cooperative condition) or by encouraging group members to be the first to make a decision (competitive condition). Crucially, participants knew that some information was unshared and that unshared information was more important than shared information for the decision process. This allowed participants to strategically withhold information, directly linking sharing and withholding information to cooperative and competitive motives, respectively. In contrast, in the current study, participants were not informed that some information was unshared and more important. Therefore, deliberately withholding information did not function as the “competitive alternative” to information sharing. We do not know what motivated people in the current study, but given that helping behavior in the Egg Hunt game measured opposing behavioral tendencies in the Hidden Profile game (as evident by the negative correlation), it might be argued that behavior in the latter reflected competitive motives. However, this is speculative and future studies should use additional measures to understand people’s underlying motives. For example, a person’s social value orientation has been shown to relate to a variety of tasks (Balliet et al., 2009; Behrens & Kret, 2019; Böckler et al., 2016; C. Boone et al., 2010). Quantified by hypothetical distributions of resources between oneself and a hypothetical partner, SVO indicates the extent to which people take into account the welfare of another person when distributing resources between oneself and that other person (van Lange, 1999). Based on the distribution, participants are classified as generally being proself or prosocial (Van Lange, De Bruin, Otten, & Joireman, 1997). Thus, this measure could help shed light on why people share information in the Hidden Profile game.

To conclude, the second component of the PCA in our study indicates that the Egg Hunt game and the Hidden Profile game measure opposing tendencies—with the former reflecting helping behavior and the latter potentially indicating a person's competitive motives.

Finally, we observed that behavior during the Tangram game was not related to behavior during any of the other games. The game is designed to measure cooperation, but possibly, methodological differences with studies using the same game might explain its lack of correlation with the other games. Cooperation was operationalized as the difference in completed puzzles between the cooperative and individual condition. While we accounted for individual differences in spatial ability necessary to complete the task within the individual condition, two issues arose from the comparison of the two conditions. First, because people were randomly paired, their spatial skills were not matched, which is particularly relevant to the performance in this game. Consequently, if a person who performed poorly individually was paired with a skilled person, the skilled individual would inevitably complete most of the puzzles in the cooperative condition. As a result, the cooperation rate is overestimated for the poorly performing person and underestimated for the skilled one. Second, for all puzzles, the difficulty level was relatively high from the beginning. Although the first puzzles included highlighted pieces that made it easier to detect their individual shapes, this might not have been enough to prevent floor effects. Moreover, in order to equalize the duration of the different games in this study, participants had less time to complete the Tangram game than in two earlier studies (Saleem, Anderson, & Barlett, 2015; Vink et al., 2018). In sum, the random matching of participants with different skills and the difficulty of the task might explain why cooperative behavior in our Tangram game was not related to the behavior in the other games. Performance during the Tangram game might therefore be a reflection of participants' spatial abilities rather than prosocial tendencies, but future studies need to verify this presumption.

In sum, we extend previous findings by showing that not only different economic games, but also changes within the same economic game elicit similar prosocial behavior. In other words, as long as the principle of the game (i.e., the pay-off structure) stays similar, methodological changes in the response scale and the way people give their response still allows researchers to measure similar prosocial tendencies by these variants. Furthermore, the aim of the current study was to investigate whether the consistent behavioral tendency observed in economic games would generalize to more naturalistic games. The short answer is: no. Our results revealed that the social dilemma games measured different behavioral tendencies than the naturalistic games. Our discussion suggests that methodological differences in, for example, the clarity on *how* to act prosocially might explain such distinction between the economic and naturalistic games. This explanation is, however, speculative and further research is needed to draw strong conclusions about what drove the current findings. Economic games are designed to create conflicts of interests within individuals and changing the payoff structure determines the nature of these conflicts. Such experimental control combined with an extensive body of literature has portrayed a rather detailed picture of the underlying motives of choosing one over another option in these games.

However, for the more naturalistic, less often used games such as the Egg Hunt and Tangram game, the motivation behind people's behavior is more ambiguous and the effects of methodological changes are less well known. This does not mean that economic games measure prosocial behavior better. In fact, we encourage researchers to include naturalistic games to investigate whether previous findings generalize to these games. However, researchers need to be aware of potential differences between games and we advise to combine naturalistic games with other paradigms to tap into potentially different aspects of prosocial behavior and to integrate the findings with existing literature. Our current study takes a first step in this direction and sheds light onto the generalizability of prosocial behavior as measured by different paradigms.

