Decision-Making in the Prisoner's Dilemma Game: The Effect of Exit on Cooperation and Social Welfare

Tessa Haesevoets[a], Dries H. Bostyn[a], Chris Reinders Folmer[b], Arne Roets[a], and Alain Van Hiel[a]

[a]Ghent University, Department of Developmental, Personality and Social Psychology, Henri Dunantlaan 2, B-9000, Ghent, Belgium

[b]Erasmus University Rotterdam, Erasmus School of Law, Burgemeester Oudlaan 50, 3062 PA, Rotterdam, The Netherlands

WORD COUNT: Xxx

Authors' Note

Declarations of interest: none. Correspondence concerning this article should be addressed to Tessa Haesevoets, Department of Developmental, Personality and Social Psychology, Henri Dunantlaan 2, B-9000, Ghent, Belgium. E-mail: Tessa.Haesevoets@UGent.be, Tel.: ++32 (0)9 264 62 69, Fax.: ++32 (0)9 264 64 99.

Abstract

The Prisoner's Dilemma Game (PDG) is a mixed-motive game that offers two players the simultaneous choice between a cooperative and a defective alternative. An often-neglected aspect of such a binary-choice game, however, is that in many real-life encounters people can choose not only to cooperate or defect, but they also have a third option: To exit the social dilemma. Although in the literature a consensus has emerged that the addition of an exit opportunity benefits cooperation, there is only scant research into its effect on social welfare. In order to allow a direct comparison of cooperation rates and welfare levels across binary-choice and trinary-choice games, in the present study we used a design in which the same participants played similar games with and without an exit option (i.e., a within-subjects design), and this in a range of structural variations. The findings of our study indicated that the aggregated outcome of both players is generally lower in games with an exit option than in games without an exit option. Moreover, our results showed that the efficiency of the exit option strongly depends on the specific outcome structure of the game (in terms of its endowment size, (a)symmetry, and level of non-correspondence). In the discussion, it is argued that the implementation of an exit option as a strategy to increase social welfare should be critically assessed.

Keywords: exit option; prisoner's dilemma game; structural variation; cooperation; social welfare

Decision-Making in the Prisoner's Dilemma Game:

The Effect of Exit on Cooperation and Social Welfare

Social dilemmas are situations in which it is more profitable for each individual to act selfishly, but such behavior harms the collective (Dawes, 1980; Messick & Brewer, 1983). Manifestations of social dilemmas are, among others, the choice whether or not to conserve community resources, to use public transportation, to donate to charity, to volunteer, and to vote in an election (Van Lange, Joireman, Parks, & Van Dijk, 2013). As these examples illustrate, social dilemmas are omnipresent and of concern to many people. In order to gain a better understanding of social dilemmas, researchers have modelled these situations into mixed-motive games in which people must choose between cooperative and defective alternatives (Dawes & Messick, 2000). In this regard, the Prisoner's Dilemma Game (PDG) is the most studied mixed-motive game, which has been used in literally hundreds of empirical studies (Au & Komorita, 2002). In this game, two persons each face a binary choice between a cooperative and a defective alternative. The relative order of the four possible outcome options defines this game. Particularly, the payoffs for the various outcomes are arranged in such a way that the best possible outcome for an individual is to unilaterally defect (DC), the second best outcome is mutual cooperation (CC), the second worst outcome is mutual defection (DD), and the worst outcome is to unilaterally cooperate (CD).

Although the PDG provides the most straightforward example of a mixed-motive game, it neglects important properties of many real-life encounters. One such property is that people's choices are often not restricted to either cooperation or defection. Indeed, many natural situations involve the opportunity to leave (or to not even enter) the interaction (for some illustrations, see Orbell & Dawes, 1991; Orbell, Schwartz-Shea, & Simmons, 1984; Seale, Arend, & Phelan, 2006). Exit options are important to investigate because they offer participants the opportunity not to be dependent upon their counterpart's choice, and as such avoid the possible losses that are associated with being exploited (see Weber, Malhotra, & Murnighan, 2004). In this vein, prior research has shown that in situations of interdependence people might have a preference to avoid the situation. Indeed, research on negotiations has shown that instead of making a risky choice, when given the opportunity, people might prefer to just exit the interaction (Shalvi et al., 2013).

Referring to the "prisoner" aspect of the PDG, Boone and Macy (1999) aptly noted that "there is surprisingly little research on what happens when the doors to the prison are unlocked" (p. 33). Interestingly, of these scant studies of how the inclusion of an exit option (besides the options to cooperate and defect) affects cooperation rates in the PDG, most used simulations that focused on the effectiveness of different strategies in iterated games with parties who are involved in multiple rounds of the game (e.g., Congleton & Vanberg, 2001; Vanberg & Congleton, 1992; Yamagishi & Hayashi, 1996). These simulations support the notion that an exit option can increase cooperation rates, because the possibility to exit allows a party to escape from dysfunctional interactions. However, based on these simulations, it is unclear how "real" participants react to an exit opportunity. In response to this, some recent studies have used an evolutionary perspective to investigate the effect of exit opportunities on cooperation in experiments with humans. In these studies, exit was operationalized as the possibility to switch partners, rather than to not be dependent on them; nevertheless, their underlying logic (one has to withdraw from an interaction in the first phase) and payoff structure are very similar to that of an exit option. In general, their results also seem to suggest that the opportunity to leave an interaction will increase cooperation rates (e.g., Barclay & Raihani, 2016; Rand, Arbesman, & Christakis, 2011).

The question whether exit also positively affects the aggregated outcome of the players (i.e., social welfare) remains, however, largely unaddressed, as only a few prior studies have investigated how exit opportunities influence the revenues of the players. A notable exception in this regard is the classic study conducted by Orbell and Dawes (1993). To minimalize the strategic effects from repeated play, participants in this study played five PDGs − each with a different interaction partner. In the binary-choice condition participants were obliged to choose between cooperation and defection, whereas in the trinary-choice condition the cooperative and defective alternatives were accompanied with an additional exit option. When one player selected the latter option, the payoff of both players was lower than the payoff of mutual cooperation, but higher than the payoff of mutual defection (cf. Seale et al., 2006; Vanberg & Congleton, 1992). Orbell and Dawes concluded that the opportunity to exit increased the aggregated outcome of the players as well as the relative gains of intended cooperators. According to the authors, these positive effects occur because cooperators are less likely to exit than defectors, causing levels of mutual cooperation to increase relative to situations where exit is not possible. Over the years, numerous scholars have cited this study (for some influential examples, see Frank, Gilovich, & Regan, 1996; Kollock, 1998; Ostrom 1998; Rand et al., 2011; Sigmund, De Silva, Traulsen, & Hauert, 2010; Tooby, Cosmides, & Price, 2006; Wilson & Dugatkin, 1997), and there has been a growing consensus in the literature that not only cooperation, but also social welfare increases when players are offered the option to leave the interaction.

The aim of the present contribution was to build further upon the work of Orbell and Dawes (1993), in order to investigate in more detail the effect of exit on cooperation and social welfare. We believe that such a follow-up study is necessary because certain aspects of Orbell and Dawes' study design might have had unintended effects on their findings. First, the availability of the exit option was manipulated between-subjects in the study of Orbell and Dawes (but also in various other exit studies). This design feature does not allow a direct comparison of the same individual's game behavior across identical binary-choice and trinary-choice games. Instead, these authors had to make inferences about the numbers of cooperators and defectors opting out by extrapolating the observed numbers in the binary-choice groups to estimated numbers of “intending cooperators” and “intending defectors” in the trinary-choice groups. In the present study, we employed a within-subjects design in which each participant was confronted with both game types. This procedure allowed a direct comparison of cooperation and defection rates in similar games with and without the option to exit.

Secondly, in the study of Orbell and Dawes (1993) the exit option was set-up in such a way that if one player selects this option, both players receive no earnings (i.e., a payoff of zero). Because of this particular set-up, unilateral defection and mutual cooperation had to generate positive earnings, whereas the earnings of mutual defection and unilateral cooperation had to be negative (i.e., DC > CC > E = 0 > DD > CD). Each game’s outcome structure thus consisted of a combination of gains and losses (for similar outcome structures, see Seale et al., 2006; Vanberg & Congleton, 1992). In our study, the relative order of the outcome options was similar as in prior exit research, but all outcome options generated positive outcomes (i.e., DC > CC > E > DD > CD = 0). We did this because prior research has indicated that gain- and loss-framed outcomes may affect decisions differently (see Aquino, Steisel, & Kay, 1992; Brewer & Kramer, 1986; McDaniel & Sistrunk, 1991).

Additionally, although Orbell and Dawes (1993) varied the payoff matrices among their five games, they did not systematically manipulate specific game parameters in order to create these matrices. In retrospect, it should be acknowledged that the matrices selected by Orbell and Dawes were skewed toward defection, as of the five games two provided a strong and three (of which one was a duplicate) a moderately strong incentive to defect. Yet, none of the included matrices provided a strong incentive to cooperate. Hence, although Orbell and Dawes included strong game variations, they heavily focused on matrices that are situated within the spectrum of defection. In order to investigate the effects of exit over a broader range of possible outcome structures, in the present study we employed matrices that are more balanced. More specifically, we included a greater number of payoff matrices in which the incentive to cooperate or defect was systematically varied through the manipulation of three crucial game parameters (i.e., endowment size, (a)symmetry, and non-correspondence), which are expected to have a significant influence on game behavior in both binary-choice and trinary-choice games.

Objectives of the present study

The present study aimed to extend prior exit research, and in particular the work of Orbell and Dawes (1993). In doing so, the present study centered on the clarification of five critical empirical issues.

1. How does the addition of an exit option influence cooperation rates?

Our first aim was to investigate in greater detail how the addition of an exit option influences cooperation rates in the PDG. Are people more likely to cooperate when the cooperative and defective options are complemented with an exit option? The results of Orbell and Dawes (1993; for similar findings, see also Barclay & Raihani, 2016; Orbell & Dawes, 1991; Orbell et al., 1984; Rand et al., 2011) indeed suggest this to be the case. Yet, it is important to note that in Orbell and Dawes' (1993) study a measure of relative cooperation was used, within dyads that *did not* exit the game (i.e., number of cooperators / [number of cooperators + number of defectors]). This measure, however, did not include the participants who *did* exit. Even though the exit option may bolster the level of cooperation in the dyads that remain in the game (which thus reflects a relative increase in cooperation), it inevitably leads to a decline in the number of cooperators in absolute terms. As such, the presence of an exit option leads to a lower share of (mutual) cooperation in absolute terms. Although this absolute decrease in cooperation seems rather obvious (as the addition of a third option to a binary-choice set almost certainly reduces the absolute number of times that the first option and second option are selected), each cooperative player who selects the exit option undermines social welfare by lowering the total revenues. A critical consequence of this absolute decrease is thus that it is expected to result in a decline (rather than a gain) in social welfare. This brings us to our first research question:

RQ 1: How does the addition of an exit option influence cooperation rates, in both relative and absolute terms?

2. Which participants are most inclined to choose the exit option?

The second aim of our study was to investigate which type of person is more inclined to select the exit option. Are cooperative individuals more inclined to choose the exit option, or is this option more likely to be preferred by defective individuals? A view on who exactly uses the exit option is of importance to understand whether trinary-choice dilemmas indeed contribute to social welfare. If the presence of an exit option leads a person who would otherwise have cooperated to exit the interaction, society is deprived of the benefit that his or her cooperation would have elicited. Conversely, if the presence of an exit option leads a person who would otherwise have defected to exit, society is spared the cost that his or her defection would have incurred.

In this regard, Orbell and Dawes’ (1993) data suggested that defectors are more prone to select the exit option than cooperators. Some earlier studies also reported that defectors exit more rapidly (e.g., see Orbell & Dawes, 1991; Orbell et al., 1984). However, Orbell and Dawes’ (1993) use of a between-subjects design (in which one group of participants played the binary-choice games, and another group of participants played the trinary-choice games) meant that this prediction could not be directly confirmed. The adoption of a within-subjects design in the present study enabled us to examine directly which type of participant most often employs the option to exit the game. More specifically, our design allowed us to record participants’ behavior in the binary-choice games in terms of how often they cooperated and defected, which could subsequently be linked directly with how often they selected the exit option in the trinary-choice games.

RQ 2: Which type of person (defectors or cooperators in binary-choice games) most often selects the exit option in trinary-choice games?

3. How does the payoff structure influence choice behavior?

A third aim of the present study was to examine if the choice to exit depends on how the outcomes of the game are structured. Although in the PDG the relative order of each outcome option is fixed, this game can be set-up so that either the incentive to cooperate or the incentive to defect increases. Such game variations are important to investigate because of the prevalence of such differences in real-world social dilemmas, and because increasing or decreasing the level of conflict within the PDG might alter people's temptation to exit.

In this vein, it must be emphasized that although Orbell and Dawes' (1993) matrices included variations in terms of the magnitude of the outcomes that can be achieved by both players and the degree in which both players' outcomes converge or diverge, the influence of these two game variations on game behavior was not systematically investigated. When endowments are high, there is more at stake than when endowments are low. As such, high endowments are expected to intensify people's preference for either cooperation or defection (i.e., people who prefer cooperation will become even more cooperatively, whereas people who prefer defection will become even more defectively). The level of non-correspondence reflects the degree of conflicting interest. The more the outcomes of the two players are opposed to each other (i.e., high non-correspondence), the more room there is for defective behavior. The more the outcomes of the two players are in line with each other (i.e., low non-correspondence), the more likely it is for cooperative behavior to occur. As a result of this, high levels of non-correspondence are expected to accentuate the incentive to defect, whereas low levels of non-correspondence are expected to accentuate the incentive to cooperate. Prior research has revealed that, under binary settings, PDGs that involve greater outcome non-correspondence indeed yield lower levels of cooperation (e.g., Molm, Takahashi, & Peterson, 2000; Parks & Hulbert, 1995).

Furthermore, although most prior research has examined PDGs with identical outcomes for both players, many social interactions are characterized by asymmetry in outcomes (for some examples, see Beckenkamp, Hennig-Schmidt, & Maier-Rigaud, 2007). Indeed, many real-life interactions entail different outcomes for each player, even if everyone who is involved acts in the same way. The concept of (a)symmetry refers to whether the outcomes of both players are identical (i.e., symmetrical) or not (i.e., asymmetrical). In the present study, we included symmetric games that entail identical payoffs for both players as well as asymmetric games that present different outcomes for each player. Note that an important feature of these asymmetric games is that they actually take away the possibility to reach equality in outcomes. Indeed, in asymmetric games one player always earns more than the other does, and this regardless of the two players' choices (so even under mutual cooperation and mutual defection the outcomes of the two players differ). As such, asymmetric games are expected to accentuate the incentive to defect, whereas symmetric games (which allow equal outcomes) are expected to accentuate the incentive to cooperate. In line with this reasoning, prior studies suggest that asymmetric outcomes result in less cooperation than symmetric outcomes under binary conditions (Croson, 1999; Lave, 1965; Sheposh & Gallo, 1973). Therefore, we decided to also manipulate (a)symmetry in our study, in addition to endowment size and level of non-correspondence (the two dimensions that were also varied by Orbell and Dawes [1993], albeit not systematically).

The present study is the first (at least to our knowledge) to include in a single study design these three game parameters. Although some prior studies have investigated the influence of these game parameters under binary-choice settings, their influence in trinary-choice games remains largely unknown. As a result of this, our study of the effects of these game parameters is explorative in nature, and we therefore refrained from formulating specific hypotheses on how these parameters are expected to differently influence choice behavior under binary-choice and trinary-choice settings. Instead, we more generally assumed that if game parameters that generally lead to high cooperation rates in binary-choice games also result in high exit rates in trinary-choice games, then the addition of an exit option to these particular parameters is likely to result in a loss (rather than a gain) of social welfare. Based upon this reasoning, we formulated the following research question:

> *RQ 3: Do the different game parameters (i.e., endowment size, (a)symmetry, and non-correspondence) have parallel effects on cooperation (under binary-choice settings) and on exit (under trinary-choice settings)?*

4. How does the exit option influence the earnings of cooperators and defectors?

Although ample studies have investigated the influence of exit on cooperation, very little attention has been provided to how exit affects the players' earnings. Therefore, our fourth aim consisted of investigating the influence of exit on the earnings of cooperators and defectors. According to Orbell and Dawes (1993; also see Orbell & Dawes, 1991) the availability of an exit option heightens the outcomes of people who tend to cooperate, whereas it generally lowers the revenues of people who are inclined to compete. Interestingly, when looking more closely at Orbell and Dawes' (1993) results, it can be inferred that these differential effects are most pronounced in matrices in which the incentive to defect is highest. Yet, it is still unclear how the addition of an exit option affects the outcomes of cooperators and defectors when games are structured in such a way that the incentive to cooperate (instead of defect) is heightened. In the present study, we therefore investigated the earnings of cooperators and defectors over a broad range of payoff matrices in which the incentive to cooperate and defect was systematically varied over the continuum of possible outcomes.

RQ 4: How does the addition of an exit option influence the individual outcomes of the two players?

5. How does the exit option influence the aggregated earnings of the players?

Relative gains and losses for cooperators and defectors are one part of the puzzle, but how does the addition of an exit option influence the net wealth of the involved parties? This brings us to the fifth and final question that we aimed to tackle: Are the aggregated outcomes of the two players higher or lower in games with or without an exit option? This question is obviously of great significance (given the importance of social welfare for sustaining social stability), but remained largely unaddressed in the literature. Based on their data, Orbell and Dawes (1993) concluded that the addition of an exit option increased the aggregated payoff of the players, even though the outcomes of intended defectors were generally lower in games with an exit option. However, it might also be the case that the impact of exit is contingent on the game's specific payoff structure, such that its effect on collective welfare is more or less prominent depending on the features of the game.

RQ 5: How does the addition of an exit option influence the aggregated outcome of the two players?

Summary

In sum, we aimed to extend previous empirical findings relating to exit, by using a more elaborated study design in which type of game (binary-choice versus trinary-choice game) was manipulated within-subjects. A crucial element of our approach is that we investigated exit effects over a broad spectrum of matrices (that reflect important variations in real-world dilemma situations), by including not only games that provide a substantial incentive to defect, but also games that encourage cooperation. Moreover, we also investigated how the installation of an exit option affects the individual outcome as well as the aggregated outcome of the players.

Method

Participants

Two hundred and nine undergraduate university students participated in the experiment in exchange for course credits. Students were invited to the laboratory in groups of 35 to 45 persons. Upon arrival, each participant was seated at a separate desk in front of a computer. Informed consent was obtained before the start of the study. Participants were first informed about the payoff structure of the PDG, and then answered three questions that probed their comprehension of the game's outcome structure. Those who were not able to answer at least two of these three check questions correctly (N = 19; 9.1%) were excluded from the analyses. As such, the final data set included 190 participants (33 men, 157 women; M_{age} = 18.62, SD = 1.82).

Procedure

During the experimental session, participants played 32 single-shot PDGs. In each of these games, participants played against a different interaction partner. Participants were not directly connected to these interaction partners during the PDGs, but at the end of the experimental session they were (manually) paired with another student and paid according to the outcome of both players' decision in one of the games. Participants were first presented with a set of 16 payoff matrices in which they had to choose between cooperation and defection (binary-choice games). Next, participants were presented with the *same* 16 payoff matrices, but now the cooperation and defection options were accompanied with an third exit option (trinary-choice games). In both set of games, the order of the matrices was randomized. The exact payoff structure of each matrix is included in Appendix A. In order to get an estimate of reliability, for both the binary-choice and the trinary-choice games Matrix 9 of Appendix A was repeated a second time; choices in these repeated games correlated relatively strongly (rs > .44, ps < .001). During the experimental session, participants also completed several individual difference measures. These data are reported in a separate manuscript that deals with how personality and situational variables influence behavioral consistency in the binary-choice PDG (see Haesevoets, Reinders Folmer, Bostyn, & Van Hiel, under review).

Manipulation of the payoff structure

In both the binary-choice and the trinary-choice games, we varied the payoff matrices of the games by manipulating three crucial game parameters that reflect prevalent differences in real-world dilemma situations. Importantly, in order to systematically investigate their influence on choice behavior, these three game parameters were manipulated independently of each other (although the effect of (a)symmetry in outcomes on either player's choice is confounded with the effect of either player's endowment size; see Appendix A). The first manipulated game parameter was endowment size. In our study, we included high and low endowments; the high endowments were always two times as large as the low endowments. The second game parameter that was manipulated related to the (a)symmetry of the payoff structure. In this regard, we distinguished between symmetric games that entail identical payoffs for both players, and asymmetric games that entail different outcomes for each player. More specifically, the asymmetric games were asymmetric combinations of the high and low endowment symmetric games (cf. Beckenkamp et al., 2007). As such, in the asymmetric games the outcomes of one player were always twice as high as those of the other player. In the high endowment games participants received the high payoff and their partner the low payoff, whereas in low endowment games these outcomes were reversed.

The final manipulated game parameter related to the degree of non-correspondence. The degree of non-correspondence can be expressed in terms of Rapoport's (1967) *K*-index, which captures the benefit of mutual cooperation over mutual defection (i.e., distance between CC and DD outcomes) relative to the benefit of exploitation over the sucker's payoff (i.e., distance between DC and CD outcomes). In order to create four different non-correspondence levels, we decreased the payoff of the mutual cooperation outcome while simultaneously increasing the payoff of the mutual defection outcome to an equivalent extent (whereas the outcomes of unilateral defection and unilateral cooperation were held constant over the different game variants; cf. Schopler et al., 2001). Critically in this regard is that our broader range of matrices fully covers, and further extends, those used by Orbell and Dawes (1993). Indeed, in their study the five different matrices were characterized by a *K*-index of respectively 0.58, 0.33, 0.60, 0.50, and 0.60 (note that their third and fifth matrix were actually duplicates), whereas in our study the low, medium low, medium high, and high non-correspondence levels were characterized by a *K*-index of respectively 0.80, 0.60, 0.40, and 0.20.

Results

1. How does the addition of an exit option influence cooperation rates?

Our first research question consisted of investigating whether the exit option increases or decreases cooperation rates, in both relative and absolute terms. Table 1 reports the number and percentage of participants that have chosen each of the choice options in both the binary-choice and the trinary-choice games. Importantly, in the trinary-choice games cooperation rates were computed in *relative* terms (in relation to only defection; so here participants who decided to exit are *not* taken into consideration) and in *absolute* terms (in relation to both defection and exit; so here exit is taken into consideration).

A comparison of the reported percentages in Table 1 reveals that, across the 16 matrices, in *relative* terms the average cooperation rates increased by 7.6% (i.e., from 62.86% in the binary-choice games to 70.43% in the trinary-choice games), while the average defection rates (of course) decreased by a similar percentage (i.e., from 37.14% in the binary-choice games to 29.57% in the trinary-choice games). However, adding a third option to a binary set will almost certainly reduce the absolute number of times that the two other options are selected. Hence, when looking at the *absolute* percentages across the 16 matrices, the average cooperation rates decreased by 25.6% (i.e., from 62.86% in the binary-choice games to 37.31% in the trinary-choice games), while the average defection rates decreased by 21.0% (from 37.14% in the binary-choice games to 16.16% in the trinary-choice games).

It can hence be concluded that, even though in relative terms cooperation rates increase, the addition of an exit option in fact reduces the absolute percentage of participants that cooperates in the game, and this even to a slightly larger extent than it does for competitive participants. Although this absolute decrease in cooperation might be self-evident (and not so interesting in itself), this decrease is nonetheless important as it might entail negative consequences for the players' aggregated outcome under trinary conditions.

2. Which participants are most inclined to choose the exit option?

Prior exit research suggested that defectors are more prone to exit than cooperators. But is this indeed the case? In this regard, our second research question consisted of investigating what type of person (defectors or cooperators in binary-choice games) is most likely to exit the game under trinary-choice settings. Important in this regard is that our use of a within-subjects design allowed us to record participants' behavior in the binary-choice games in terms of how often they cooperated and defected, which could subsequently be linked directly with how often they selected the exit option in the trinary-choice games.

In order to investigate this particular research question, we first performed a *k*-means cluster analysis on participants' responses in the 16 binary-choice games, using the Akaike's Information Criterion (AIC) to determine the appropriate amount of clusters. Prior research implicitly assumed that under binary-choice conditions there are consistent cooperators and consistent defectors. However, the AIC plot in Figure 1 shows that a three-cluster solution fitted the data best (and hence better than a two-cluster solution). To interpret these three clusters we computed a cooperation index by counting the total number of cooperative choices across the 16 binary-choice games. Relating the clusters to this cooperation index revealed that there is a first group of participants that strongly prefers defection ($M_{CoopIndex}$ = 3.36, *SD* = 2.38; *N* = 44), a second group of participants that switches between defection and cooperation ($M_{CoopIndex}$ = 9.46, *SD* = 1.86; *N* = 72), and a third group of participants that has a strongly preference for cooperation ($M_{CoopIndex}$ = 14.62, *SD* = 1.57; *N* = 74). Note that this categorization thus highlights the interesting category of switchers (in addition to defectors and cooperators), who are actually inconsistent in their game choices.

To investigate which of these three groups of participants (defectors, switchers, or cooperators) is most prone to select the exit option (when this option is made available), we subsequently conducted an ANOVA analysis with the exit index (which reflects the total number of exit choices across the 16 trinary-choice games) as the dependent variable and the three clusters as a predictor. This analysis revealed a significant main effect of our three groups on exit behavior, $F(2, 187) = 15.91$, $p < .001$, $R^2 = .15$. Pairwise comparisons using Tukey-tests revealed that the group of defectors ($M_{ExitIndex}$ = 9.48, *SD* = 4.02) and the group of switchers ($M_{ExitIndex}$ = 8.01, *SD* = 3.72) exited significantly more often (both *p*s $\leq .001$) than the group of cooperators ($M_{ExitIndex}$ = 5.47, *SD* = 4.73). The difference between defectors and switchers, however, did not reach statistical significance ($p = .063$).

In sum, the results of our analyses showed that people who often defect in binary-choice games more often select the exit option in trinary-choice games than people who frequently cooperate in binary-choice games. Based on this finding, it can be concluded that defectors are indeed more prone to exit than cooperators. However, our analyses additionally revealed that exit also serves as an escape road for those participants who feel no strong tendency for either cooperation or defection in the binary-choice games (i.e., the group of switchers, who seem to be indecisive in their choices for either cooperation or defection).

3. How does the payoff structure influence choice behavior?

Our third research question pertained to the effects of game outcome structure on choice behavior in both binary-choice and trinary-choice games. More specifically, we examined if the systematic manipulation of three crucial game parameters – that is, endowment size, (a)symmetry, and level of non-correspondence – has similar effects on cooperation (in the binary-choice games) and exit (in the trinary-choice games). Note that such a structural investigation of the different game features is important in order to be able to judge the efficacy of the exit option. Whereas the influence of our three game parameters on choice behavior in the binary-choice games was investigated within a binomial framework, the analysis of the trinary-choice games required a multinomial approach. The results of both types of analyses are described below.

Influence of payoff-structure under binary-choice conditions

We first examined the data of the binary-choice games with a generalized linear mixed model (GLMM), estimating both random intercepts and random slopes at the participant level, with a logistic link function using the lme4 package in R (Bates, Maechler, Bolker, & Walker, 2015; R Core Team, 2017). The statistical results (odds ratios, standard errors, and *p*-values) of these analyses are reported in Table 2. This table reports the main effect of endowment size, (a)symmetry, and non-correspondence, as well as the interaction effect between endowment size and (a)symmetry. We incorporated this particular interaction because it is an essential part of our design (i.e., in order to create asymmetric outcomes we used the two different endowment sizes). To ease the interpretation of our results, below we discuss the effects of these game parameters in terms of how they influenced the predicted probabilities for participants' decisions. The predicted probabilities are visualized in Figures 2a to 5a.

Several interesting results emerged from these analyses. First of all, our model showed a significant main effect of endowment size. This effect is displayed in Figure 2a, which shows that in binary-choice games participants were more likely to cooperate in games with high rather than low endowments. Additionally, our model also revealed a significant main effect of (a)symmetry. As displayed in Figure 3a, participants were more likely to cooperate in symmetric binary-choice games than in asymmetric binary-choice games. However, the results for endowment size and (a)symmetry were also qualified by a significant interaction, which is visualized in Figure 4a. This figure shows that in asymmetric binary-choice games participants were more likely to cooperate under high than under low endowments. In symmetric binary-choice games, however, endowment size hardly affected cooperation rates. Finally, our model also revealed a significant main effect of non-correspondence. As shown in Figure 5a, the predicted probability of cooperation was higher in binary-choice games with lower levels of non-correspondence than in binary-choice games with higher levels of non-correspondence.

Influence of payoff-structure under trinary-choice conditions

To analyze the influence of the three game parameters on choice behavior in the trinary-choice games, we subsequently fitted a multinomial mixed logit model using the mlogit package in R (Croissant, 2012). We report the results of a random intercept-only analysis (as random slope models did not fully converge). The statistical results (odds ratios, standard errors, and *p*-values) of these analyses are reported in Table 3. Again, this table reports the main effect of the three situational manipulations, plus the interaction between endowment size and (a)symmetry. Note that in these analyses the defection option was used as the reference category, so in Table 3 the odds ratios for cooperation and exit are given with respect to the defection option. Because this makes our results very hard to interpret, below we again discuss our results in terms of predicted probabilities. Figures 2b to 5b display the predicted probabilities.

Again, several interesting findings emerged. The main effect of endowment size is shown in Figure 2b. This figure reveals that in the in trinary-choice games the predicted probability of cooperation was very similar in games with high and low endowments, but the predicted probability of exit behavior was higher in games with high endowments than in games with low endowments. As a result of this, the probability of defection was lower in high endowment trinary-choice games than in low endowment trinary-choice games. We uncovered a similar pattern of results for symmetric versus asymmetric trinary-choice games (see Figure 3b): The (a)symmetry manipulation did not seem to influence participants' rates of cooperation, but symmetric games were associated with higher levels of exit behavior and lower levels of defection than asymmetric games. Here too, endowment size and (a)symmetry interacted significantly. This interaction effect is shown in Figure 4b. In asymmetric trinary-choice games high endowments caused higher rates of exit behavior and lower rates of defection than low endowments (while cooperation stayed almost the same), whereas endowment size hardly had an effect on participants' choices in symmetric trinary-choice games. Finally, the main effect of non-correspondence is shown in Figure 5b. This figure reveals that, in trinary-choice games, lower levels of non-correspondence were associated with higher rates of cooperation and lower rates of both defection and exit behavior than higher levels of non-correspondence.

Summary

From these analyses, it can be concluded that in both the binary-choice and the trinary-choice games participants' choices were significantly affected by the structure of the game. And, even more importantly, a comparison of Figure 2a with Figure 2b and of Figure 3a with Figure 3b reveals that those game parameters that most strongly encourage cooperation in the binary-choice games (i.e., high endowments and symmetry in outcomes) have almost no effect on cooperation in the trinary-choice games. Instead, in the trinary-choice games these particular parameters most strongly boost the predicted probability of exit behavior. The parallel effects of endowment size and (a)symmetry on cooperation (in the binary-choice games) and exit (in the trinary-choice games) are of special importance because they may have repercussions for the outcomes that the different game variants generate. For the non-correspondence manipulation, we did not find such parallel effects (compare Figure 5a with Figure 5b).

4. How does the exit option influence the earnings of cooperators and defectors?

Our fourth research question consisted of investigating how the exit option affects the players' individual outcomes. In order to answer this particular question, for each of the 16 games, we computed the earnings of cooperators and defectors (and this separately for the binary-choice and the trinary-choice games). Table 4 provides an overview of these individual earnings. Note that these calculations are based on the reported frequencies of the different outcome options within the entire study population and on the payoff that each outcome option generates (see Table 1 for the exact percentages and Appendix A for the exact payoff structure of each matrix). A more detailed description of these calculations is provided in Appendix B.

As shown in Table 4, in 11 of the 16 matrices cooperators obtained higher outcomes in games with an exit option than in games without an exit option. This is also reflected by the finding that, across the 16 matrices, cooperators earned about 10.5 euro more in the trinary-choice games than in the binary-choice games (i.e., a total of 115.97 euro in the binary-choice games versus 126.45 euro in the trinary-choice games), which is thus equivalent with a gain of 9.2% − a difference that is statistically significant, $t(15) = 2.40$, $p = .03$. Notably, the five matrices in which the outcomes are lower with exit than without exit are the ones in which the higher cooperation rates were reported in the binary-choice games (compare the earnings reported in Table 4 with the percentages reported in Table 1). Conversely, in 13 of the 16 matrices defectors earned less in games with an exit option than in games without an exit option. Across al 16 matrices, defectors earned about 19.1 euro (8.9%) less in the trinary-choice games (a total of 156.91 euro) than in the binary-choice games (a total of 175.97 euro), $t(15) = -4.57$, $p < .001$.

Taken these findings together, it can be concluded that across all games the addition of an exit option led to an increase in the earnings of cooperators, but at the same time it decreased the earnings of defectors. The opportunity to exit the game hence thus results in gains for cooperators, but losses for defectors.

5. How does the exit option influence the aggregated earnings of the players?

In order to investigate our fifth research question, which asks how the exit option affects social welfare, for each matrix we also computed the aggregated outcome of the players. Table 4 provides an overview of the players' aggregated earnings. Here too, these calculations are based on the frequencies of the different choice options within the entire study population and on the payoff of each of the possible outcomes (see Appendix B for more information on these calculations).

Table 4 shows that in 10 of the 16 matrices the aggregated outcome of the two players was lower when the option to exit was present. This was also reflected by a loss of about 20.5 euro across the 16 matrices in the trinary-choice games relative to the binary-choice games (i.e., a total of 276.65 euro in the binary-choice games versus 256.20 euro in the trinary-choice games). The aggregated earnings in the trinary-choice games were thus 9.3% lower than in the binary-choice games, this reflects a significant difference, $t(15) = -2.77$, $p = .014$.

Critically, a closer inspection of Table 4 (in comparison with the percentages reported in Table 1), however, reveals that whether exit increases or decreases the aggregated outcome of the players strongly depends on the game's outcome structure, and more specifically the extent to which the game evokes cooperation or defection. Indeed, the ten matrices in which the addition of an exit option reduced the aggregated outcome are actually those that induced the higher cooperation rates (more than 62% cooperation; see Table 1) in the binary-choice variants. Conversely, the six matrices in which the addition of an exit option slightly enhanced the aggregated outcome of the players (although it must be stressed that these positive exit effects were very small and almost trivial) are the ones that induced the higher defection rates (more than 46% defection; see Table 1) in the binary-choice variants.

It can hence be concluded that, when taking a broad range of payoff matrices into account, the addition of an exit option decreased the aggregated outcomes of the players (i.e., the aggregated outcome of the two players is lower in games with an exit option than in games without the possibility to exit). Moreover, our results also indicate that the addition of an exit option seems to be especially negative for those games that evoke higher cooperation rates in the binary-choice game variants.

Discussion

Many societal conflicts arise from competing interests in terms of concerns for oneself versus the collective. A wide range of mixed-motive games has been employed in research to study these competing interests, with the most studied game being the PDG that provides people the choice between cooperating and defecting. However, a neglected aspect of this game is that in many real-life encounters people are not obliged to choose either option, as they can often also decide to leave the interaction. The present study builds further on prior exit research by Orbell and Dawes (1993; also see Barclay & Raihani, 2016; Congleton & Vanberg, 2001; Orbell & Dawes, 1991; Orbell et al., 1984; Rand et al., 2011; Seale et al., 2006; Vanberg & Congleton, 1992; Yamagishi & Hayashi, 1996), which has resulted in a consensus in the literature that exit benefits cooperation and, as a result, social welfare. Scholars have, for instance, noted exit research to demonstrate “that defectors were more likely to exit than were cooperators” (Boone & Macy, 1999, p. 34), “higher cooperation rates with an exit option than in the standard PD" (Bohnet & Kübler, 2005, p. 63), "that cooperators ended up with greater wealth at the end of play than non-cooperators" (Boone & Buck, 2003, p. 170), and "that when players are free to accept or reject play in PD-games, the aggregate social welfare increases" (Kurrild-Klitgaard, 2010, p. 349).

The major purpose of the present research was to explicitly test these assumptions with an expanded study design in which the same participants played similar PDGs with and without the option to exit, and in which structural game features were varied to explore exit effects in a more systematic way. This procedure also allowed us to explore the efficiency of the exit option as a strategy to increase the welfare of society. More specifically, we aimed to provide an answer to the following research questions: (1) how the addition of an exit option influences cooperation rates, (2) which participants are most likely to choose the exit option, (3) whether the choice to exit is dependent upon the structure of the game, and how the addition of an exit option influences (4) the earnings that cooperators and defectors obtain and (5) the aggregated outcome of the two players.

Main conclusions

The following five conclusions can be drawn from our study. First, with regard to our research question how exit influence cooperation rates, our results indicated that although the addition of the option to exit increased cooperation rates in relative terms, it nevertheless lowered both cooperation and defection rates in absolute terms. And, particularly interesting in this regard is our finding that the absolute decrease in cooperation is somewhat more pronounced than the absolute decrease in defection. Even though this absolute decrease in cooperation may not be important on its own, the importance of this finding lies in the fact that it has negative consequences for social welfare (see below). Secondly, in light of the research question which participants are most likely to exit, our data showed that not only defectors, but also participants who switch between cooperation and defection in binary-choice games were more prone than cooperators to employ the exit option in trinary-choice dilemmas. In light of our third research question, we examined the influence of game structure on choice behavior. Our analysis indicated that the game's outcome structure has a significant influence on exit behavior. Most importantly, however, is the finding that two of the three parameters that most strongly encourage cooperation in binary-choice situations (i.e., high endowments and symmetric games) also seem to most strongly favor exit in trinary-choice situations.

Our two final research questions relate to the outcomes that the two game types generate. In this regard, our fourth research question focused on how exit influences outcomes of cooperators and defectors. The present study indicated that although across all games the addition of an exit option led to an increase in the earnings of cooperators, it decreased the earnings of defectors. Hence, in the presence of an exit option cooperators individually earned more, defectors less. Our fifth and final research question consisted of investigating the influence of exit on social welfare. Even though exit benefited cooperation within a selected group of players that remained in the game (which reflects a relative increase in cooperation), the absolute decrease in cooperation that resulted from adding an exit opportunity nonetheless negatively affected social welfare. Particularly, our data revealed that the aggregated outcome of the two players was almost 10% lower in games with an exit option than in games without the possibility to exit. And, this negative effect of exit on collective revenues was strongest in games that were set-up to encourage cooperation.

How do our findings relate to prior exit studies?

If we compare the results of our study with prior exit research, we note some consistencies but also some notable differences. A first notable difference is that in our study cooperation rates under binary-choice conditions were higher than in the study of Orbell and Dawes (1993). More specifically, Orbell and Dawes reported cooperation rates that varied between 35% and 54% in their binary-choice games (see their Table 4, p. 794), whereas in the present study these rates varied between 40% and 80% (see our Table 1). A first possible explanation for this difference is that in Orbell and Dawes none of the included matrices provided a strong incentive to cooperate, whereas we included both matrices that provide a strong incentive to cooperate as well as matrices that provide a strong incentive to defect. A second possible explanation for this difference, however, is that the participants in our sample were more prosocial than the ones in the study of Orbell and Dawes.

Secondly, our study corroborates prior evidence that, relative to defection, cooperation increased in the presence of an exit option (see Orbell & Dawes, 1993; see also Barclay & Raihani, 2016; Orbell & Dawes, 1991; Rand et al., 2011). Yet, even though cooperation relative to defection increased, adding a third option (i.e., exit) to two behavioral choices (i.e., cooperation and defection) inevitably leads to a decrease of the absolute number of participants that cooperated, which leads to a loss of social welfare. By focusing exclusively on relative figures of cooperators versus competitors in prior studies, previous research appears to overlook the notion that exit can lead to welfare losses and therefore is a suboptimal strategy to maximize outcomes.

Several previous studies (e.g., Orbell & Dawes, 1991, 1993; Orbell et al., 1984) have reported that defectors were more prone to exit than cooperators. However, this prior research could not directly infer which individuals were most likely to exit, as different groups of participants engaged in the binary-choice and the trinary-choice games. In the present study, we compared decision behavior of the same participants in binary and trinary situations by using a within-subjects design. This direct comparison complements prior exit research by showing that not only defectors, but also switchers often opt to exit when this is possible. These switchers are an interesting and fairly large group of participants that is expected to be indecisive when presented with the choice between cooperation and defection, and therefore may prefer not to make a decision between cooperation and defection when there is a possibility to exit the interaction. Note that our finding that switchers prefer exit above cooperation and defection is interesting, as it can also help us explain why not only defection, but also cooperation rates substantially decline when introducing an exit opportunity. Although the present study clarifies who is most prone to exit, it is still unclear why they do so. Future research should tackle this “why” question.

Most prior exit studies did not investigate how the outcome structure of the game influences participants’ choices, although differences in the outcome structure are prevalent in real-life settings. A remarkable exception is the seminal study by Orbell and Dawes (1993), which investigated, in a rather unsystematic way, how variations in endowment size and level of non-correspondence influence choice behavior. When we look across our different game variants, the addition of an exit option increased the relative welfare of cooperators, but at the same time decreased that of defectors. In this regard, our findings are actually very similar to those of Orbell and Dawes (1993; also see Orbell & Dawes, 1991), who also reported gains for intended cooperators and losses for intended defectors. However, when looking at the aggregated outcomes of the players, our data revealed that exit negatively affects social welfare. How can this latter finding be reconciled with the conclusion of Orbell and Dawes (1993) that social welfare increases when exit is available (see their Table 3, p. 793)? Importantly, these authors only included matrices in which the incentive to defect was rather high. We, on the other hand, included a broader range of payoff matrices in which the incentive to cooperate and defect was systematically varied. In this regard, we particularly selected our matrices in such a way that they covered almost the whole spectrum of non-correspondence (i.e., the *K*-indices of our matrices varied from 0.20 to 0.80). An interesting feature of our study is that our range of outcomes covers *and* extends the range of outcomes investigated by Orbell and Dawes (1993). In this light, our findings extend prior exit research by showing that, if in addition to games that promote defection also games that promote cooperation are included, exit actually has an overall negative impact on social welfare. These findings also illustrate that Orbell and Dawes' (1993) selection of particular matrices might unintentionally have steered their conclusions, which underscores the great importance of considering various structural game features when investigating social dilemmas empirically.

Limitations, Implications, and directions for future research

A limitation of the present study is that we did not counterbalance the presentation order of the binary and the trinary set of games. While the order of games within either set was fully randomized, all participants were first presented with the set of binary situations, followed by the set of trinary situations. As a result of this, it cannot be ruled out that certain differences between the two games types might be due to the order in which they were presented. In order to rule out possible order effects, future studies should counterbalance the order of the different choice sets. Despite this particular methodological limitation, our findings rather consistently revealed that the addition of an exit opportunity mostly had a negative effect on social welfare when the dilemma situation was set-up in such a way that it elicited high cooperation rates. As such, a first important implication that arises from our findings is thus that, particularly in situations in which cooperation is the dominant response, exit options should not be implemented, as they may lower, rather than increase, the collective revenue. This implication may be especially relevant to policy makers in their attempts to enhance cooperation in real-world dilemma situations.

Another interesting implication concerns the notion that the option to exit may promote the departure of not only defectors, but also of people who switch between cooperation and defection. The notion that these individuals may prefer exit above defection implies that their competitive behavior in binary situations may not only reflect the desire to exploit others, but also to shield oneself from their anticipated competitiveness (i.e., aversive competition; see Ten Velden, Beersma, & De Dreu, 2009). Importantly, research has demonstrated that aversive competitors display substantial cooperation when presented with a cooperative partner, but chose to defect in the absence of such guarantees. Moreover, our findings on switchers imply that exit options also promote the departure of conditional cooperators, who may also exit out of fear of exploitation, despite preferring mutual cooperation. Accordingly, to understand how to maximize the conciliatory potential of trinary-choice situations, it is crucial to learn how exit may be avoided, particularly in aversive competitors and conditional cooperators. To do so, future research might, for instance, explore the effectiveness of communication and commitment devices, which have been shown to effectively promote cooperation in binary settings (Orbell, Van de Kragt, & Dawes, 1988).

Ample prior decision-making research has shown that when games are played repeatedly with the same person, one player’s choices considerably influence the other player’s choices (e.g., Baker & Rachlin, 2001; Silverstein, Cross, Brown, & Rachlin, 1998). In this light, another interesting implication concerns the notion how the observed processes may develop over a longer timeframe. In single-shot interactions, the addition of an exit option not only rewards cooperators, but at the same time it penalizes defectors. An important question for future research is how these outcomes may influence decision makers in the longer run. Especially interesting is the question whether exit eventually will encourage particularly defectors to change their behavior − by reducing their tendency to defect in favor of exit, or even by adopting more cooperative patterns of interaction. Theoretical work in evolutionary biology suggests that exit options (or “optional games”) may evoke complex patterns of interaction over longer time periods, as their positive effect on cooperative interactions may eventually create circumstances that are ideal for an invasion by defectors (see Batali & Kitcher, 1995). Such simulations, however, tend to focus on the survival rates of particular strategies, rather than on the way rationally bounded individuals may switch between choices in response to their social environment (or expectations thereof). As such, further insight into the “psycho-logic” of these processes in human players may prove valuable. For instance, reputation may operate differently under trinary than under binary conditions. On the one hand, shunning possible defectors may involve lower costs to players’ cooperative reputation than reciprocating (see Panchanathan & Boyd, 2004), and favorable reputations may help to reassure conditional cooperators (see previous point). On the other hand, however, shunning may represent a less powerful punishment to defectors, and may lead negative reputations to be forgotten rather than perpetuated through non-cooperative play. As these examples illustrate, studies into the “psycho-logic” of trinary games may enlarge our understanding of why and when people opt to exit.

On a related note, the present results are also interesting in light of recent research by Van de Calseyde, Keren, and Zeelenberg (2017) which illustrated that buying insurance against the risk of betrayal has a hidden cost. More specifically, their findings revealed that the use of financial safeguards, which are intended to minimize the risky nature of acting cooperatively, might paradoxically increase the probability of betrayal (also see Falk & Kosfeld, 2006; Malhorta & Murninghan, 2002, for similar findings). In our study, choosing the exit option resulted in a fixed intermediate outcome that is located in between the mutual cooperation and the mutual defection outcome, and can therefore also be interpreted as some sort of insurance against defection by the other player. Because choosing the exit option (implicitly) signals distrust, it can be expected that choosing exit may also have negative social consequences in ongoing interactions. That is, it is possible that people will show less cooperation when confronted again with someone who previously has chosen to exit the interaction. In the long-run, the act of exiting on itself can thus be expected to undermine future cooperation (and as a consequence also social welfare). Future scholars who conduct exit studies are therefore strongly encouraged to investigate how exit operates when people interact with others in iterated games.

Conclusion

Does the presence of an exit option benefit cooperation and welfare? Often it does not. Our findings indicate that, under certain circumstances, exit is a suboptimal strategy to maximize social welfare because it lowers the absolute number of cooperators and as a result also the collective revenue. Moreover, our research highlights that when investigating exit effects it is of crucial importance to take structural elements of the situation into account, as these have a substantial impact on the efficacy of the exit option. More generally, the present study showed that both research and practice can gain greatly in richness by giving more consideration to exit options in the study of cooperation, and we wholeheartedly encourage researchers to follow in Orbell and Dawes' (and others) footsteps by doing so.

References

Aquino, K., Steisel, V., & Kay, A. (1992). The effects of resource distribution, voice, and decision framing on the provision of public goods. *Journal of Conflict Resolution, 36,* 665-687.

Au, W. T., & Komorita, S. S. (2002). Effects of initial choices in the prisoner's dilemma. *Journal of Behavioral Decision Making, 15,* 343-359.

Baker, F., & Rachlin, H. (2001). Probability of reciprocation in repeated prisoner's dilemma games. *Journal of Behavioral Decision Making*, *14*, 51-67.

Barclay, P., & Raihani, N. (2016). Partner choice versus punishment in human Prisoner's Dilemmas. *Evolution and Human Behavior, 37,* 263-271.

Batali, J., & Kitcher, P. (1995). Evolution of altruism in optional and compulsory games. *Journal of Theoretical Biology*, *175*, 161-171.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67, 1-48.

Beckenkamp, M., Hennig-Schmidt, H., & Maier-Rigaud, F. P. (2007). Cooperation in symmetric and asymmetric prisoner's dilemma games. *MPI Collective Goods Preprint*.

Bohnet, I., & Kübler, D. (2005). Compensating the cooperators: is sorting in the prisoner's dilemma possible? *Journal of Economic Behavior & Organization*, *56,* 61-76.

Boone, R. T., & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior, 27,* 163-182.

Boone, R. T., & Macy, M. W. (1999). Unlocking the doors of the prisoner's dilemma: Dependence, selectivity, and cooperation. *Social Psychology Quarterly*, 32-52.

Brewer, M. B., & Kramer, R. M. (1986). Choice behavior in social dilemmas: Effects of social identity, group size, and decision framing. *Journal of Personality and Social Psychology, 50,* 543-549.

Congleton, R. D., & Vanberg, V. J. (2001). Help, harm or avoid? On the personal advantage of dispositions to cooperate and punish in multilateral PD games with exit. *Journal of Economic Behavior & Organization*, *44*, 145-167.

Croissant, Y. (2012). Estimation of multinomial logit models in R: The mlogit Packages. *R package version 0.2-4*. URL: http://cran. r-project. org/web/packages/mlogit/vignettes/mlogit.pdf.

Croson, R. T. A. (1999). The disjunction effect and reason-based choice in games. *Organizational Behavior and Human Decision Processes, 80,* 118-133.

Dawes, R. M. (1980). Social dilemmas. *Annual Review of Psychology, 31,* 169-193.

Dawes, R. M, & Messick, D. M. (2000). Social dilemmas. *International Journal of Psychology*, *35*, 111-116.

Falk, A., & Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, *96*, 1611-1630.

Frank, R. H., Gilovich, T. D., & Regan, D. T. (1996). Do economists make bad citizens? *The Journal of Economic Perspectives, 10,* 187-192.

Haesevoets, T. Reinders Folmer, C., Bostyn, D., & Van Hiel, A. Behavioral consistency within various Prisoner's Dilemma Games: The role of personality and situation. *Manuscript "under review" at the European Journal of Personality*.

Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology, 24,* 183-214.

Kurrild-Klitgaard, P. (2010). Exit, collective action and polycentric political systems. *Public Choice*, *143*, 339-352.

Lave, L. B. (1965). Factors affecting co-operation in the prisoner's dilemma. *Behavioral Science, 10,* 26-38.

Malhotra, D., & Murnighan, J. K. (2002). The effects of contracts on interpersonal trust. *Administrative Science Quarterly*, *47*, 534-559.

McDaniel, W. C., & Sistrunk, F. (1991). Management dilemmas and decisions: Impact of framing and anticipated responses. *Journal of Conflict Resolution, 35,* 21-42.

Messick, D. M., & Brewer, K. (1983). Solving social dilemmas: A Review. In L.Wheeler, & P. Shaver (Eds.), *Review of personality and social psychology.* Beverly Hills, CA: Sage Publications, Inc.

Molm, L. D., Takahashi, N., & Peterson, G. (2000). Risk and trust in social exchange: An experimental test of a classical proposition. *American Journal of Sociology,* 1396-1427.

Orbell, J. M., & Dawes, R. M. (1991). A “Cognitive Miser” Theory of Cooperators Advantage. *American Political Science Review, 85*, 515-528.

Orbell, J. M., & Dawes, R. M. (1993). Social welfare, cooperators' advantage, and the option of not playing the game. *American sociological review*, 787-800.

Orbell, J. M., Schwartz-Shea, P., & Simmons, R. T. (1984). Do cooperators exit more readily than defectors?. *American Political Science Review, 78*, 147-162.

Orbell, J. M., Van de Kragt, A. J. C., & Dawes, R. M. (1988). Explaining discussion-induced cooperation. *Journal of Personality and Social Psychology, 54,* 811-819.

Ostrom, E. (1998). A behavioral approach to the rational choice theory of collective action: Presidential address, American Political Science Association, 1997. *American Political Science Review, 92*, 1-22.

Panchanathan, K., & Boyd, R. (2004). Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature, 432*, 499-502.

Parks, C. D., & Hulbert, L. G. (1995). High and low trusters' responses to fear in a payoff matrix. *Journal of Conflict Resolution*, *39*, 718-730.

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rand, D. G., Arbesman, S., & Christakis, N. A. (2011). Dynamic social networks promote cooperation in experiments with humans. *Proceedings of the National Academy of Sciences*, *108*, 19193-19198.

Rapoport, A. (1967). A note on the "index of cooperation" for Prisoner's Dilemma. *Journal of Conflict Resolution*, *11*, 101-103.

Shalvi, S., Reijseger, G., Handgraaf, M. J., Appelt, K. C., Femke, S., Giacomantonio, M., & De Dreu, C. K. (2013). Pay to walk away: Prevention buyers prefer to avoid negotiation. *Journal of Economic Psychology*, *38*, 40-49.

Schopler, J., Insko, C. A., Wieselquist, J., Pemberton, M., Witcher, B., Kozar, R., ... & Wildschut, T. (2001). When groups are more competitive than individuals: The domain of the discontinuity effect. *Journal of Personality and Social Psychology*, *80*, 632-644.

Seale, D. A., Arend, R. J., & Phelan, S. (2006). Modeling alliance activity: Opportunity cost effects and manipulations in an iterated prisoner's dilemma with exit option. *Organizational Behavior and Human Decision Processes*, *100*, 60-75.

Sheposh, J. P., & Gallo, P. S. (1973). Asymmetry of payoff structure and cooperative behavior in the prisoner's dilemma game. *Journal of Conflict Resolution, 17,* 321-333.

Sigmund, K., De Silva, H., Traulsen, A., & Hauert, C. (2010). Social learning promotes institutions for governing the commons. *Nature*, *466*, 861-863.

Silverstein, A., Cross, D., Brown, J., & Rachlin, H. (1998). Prior experience and patterning in a prisoner's dilemma game. *Journal of Behavioral Decision Making*, *11*, 123-138

Ten Velden, F. S., Beersma, B., & De Dreu, C. K. (2009). Goal expectations meet regulatory focus: How appetitive and aversive competition influence negotiation. *Social Cognition*, *27*, 437-454.

Tooby, J., Cosmides, L., & Price, M. E. (2006). Cognitive adaptations for n-person exchange: the evolutionary roots of organizational behavior. *Managerial and Decision Economics*, *27*, 103-129.

Vanberg, V. J., & Congleton, R. D. (1992). Rationality, morality, and exit. *American Political Science Review*, *86*, 418-431.

Van de Calseyde, P. P., Keren, G., & Zeelenberg, M. (2017). The hidden cost of insurance on cooperation. *Journal of Behavioral Decision Making*, *30*, 1182-1192.

Van Lange, P. A. M., Joireman, J., Parks, C. D., & Van Dijk, E. (2013). The psychology of social dilemmas: A review. *Organizational Behavior and Human Decision Processes, 120*, 125-141.

Weber, J. M., Malhotra, D., & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in organizational behavior*, *26*, 75-101.

Wilson, D. S., & Dugatkin, L. A. (1997). Group selection and assortative interactions. *The American Naturalist, 149*, 336-351.

Yamagishi, T., & Hayashi, N. (1996). Selective play: Social embeddedness of social dilemmas. In *Frontiers in social dilemmas research* (pp. 363-384). Springer Berlin Heidelberg.

Table 1.

Number and percentage of participants who choose each of the choice options in the binary-choice and the trinary-choice games.

Matrix #	Binary-choice		Trinary-choice				
	Cooperation vs. Defection		Cooperation vs. Defection (= Relative Percentages)		Cooperation vs. Defection vs. Exit (= Absolute Percentages)		
	Cooperation	Defection	Cooperation	Defection	Cooperation	Defection	Exit
1. High endowment, symmetric, low non-correspondence	148 (77.9%)	42 (22.1%)	90 (86.5%)	14 (13.5%)	90 (47.4%)	14 (7.4%)	86 (45.3%)
2. Low endowment, symmetric, low non-correspondence	151 (79.5%)	39 (20.5%)	89 (85.6%)	15 (14.4%)	89 (46.8%)	15 (7.9%)	86 (45.3%)
3. High endowment, asymmetric, low non-correspondence	128 (67.4%)	62 (32.6%)	79 (78.2%)	22 (21.8%)	79 (41.6%)	22 (11.6%)	89 (46.8%)
4. Low endowment, asymmetric, low non-correspondence	102 (53.7%)	88 (46.3%)	74 (60.2%)	49 (39.8%)	74 (38.9%)	49 (25.8%)	67 (35.3%)
5. High endowment, symmetric, medium low non-correspondence	143 (75.3%)	47 (24.7%)	79 (85.9%)	13 (14.1%)	79 (41.6%)	13 (6.8%)	98 (51.6%)
6. Low endowment, symmetric, medium low non-correspondence	150 (78.9%)	40 (21.1%)	86 (82.7%)	18 (17.3%)	86 (45.3%)	18 (9.5%)	86 (45.3%)
7. High endowment, asymmetric, medium low non-correspondence	113 (59.5%)	77 (40.5%)	82 (78.8%)	22 (21.2%)	82 (43.2%)	22 (11.6%)	86 (45.3%)
8. Low endowment, asymmetric, medium low non-correspondence	95 (50.0%)	95 (50.0%)	75 (61.0%)	48 (39.0%)	75 (39.5%)	48 (25.3%)	67 (35.3%)
9. High endowment, symmetric, medium high non-correspondence	140 (73.7%)	50 (26.3%)	67 (79.8%)	17 (20.2%)	67 (35.3%)	17 (8.9%)	106 (55.8%)
10. Low endowment, symmetric, medium high non-correspondence	142 (74.7%)	48 (25.3%)	76 (77.6%)	22 (22.4%)	76 (40.0%)	22 (11.6%)	92 (48.4%)
11. High endowment, asymmetric, medium high non-correspondence	100 (52.6%)	90 (47.4%)	64 (63.4%)	37 (36.6%)	64 (33.7%)	37 (19.5%)	89 (46.8%)
12. Low endowment, asymmetric, medium high non-correspondence	89 (46.8%)	101 (53.2%)	60 (48.8%)	63 (51.2%)	60 (31.6%)	63 (33.2%)	67 (35.3%)
13. High endowment, symmetric, high non-correspondence	123 (64.7%)	67 (35.3%)	55 (70.5%)	23 (29.5%)	55 (28.9%)	23 (12.1%)	112 (58.9%)
14. Low endowment, symmetric, high non-correspondence	119 (62.7%)	71 (37.4%)	50 (59.5%)	34 (40.5%)	50 (26.3%)	34 (17.9%)	106 (55.8%)
15. High endowment, asymmetric, high non-correspondence	93 (48.9%)	97 (51.1%)	61 (64.9%)	33 (35.1%)	61 (32.1%)	33 (17.4%)	96 (50.5%)
16. Low endowment, asymmetric, high non-correspondence	75 (39.5%)	115 (60.5%)	47 (43.5%)	61 (56.5%)	47 (24.7%)	61 (32.1%)	82 (43.2%)
Average (Matrix 1-16)	**62.86%**	**37.14%**	**70.43%**	**29.57%**	**37.31%**	**16.16%**	**46.56%**

Note. The exit option is not included in the calculation of the relative percentages for the trinary-choice games. Small deviations are due to rounding

Table 2.

Results of generalized linear mixed model for the binary-choice games.

Fixed Effects	$\hat{\beta}$	se	*z*-value	*p*
Intercept: Coop vs. Defect	-0.13	0.20	-0.66	.391
Endowment (Low): Coop vs. Defect	-0.65	0.13	-4.93	<.001
Symmetry (Symmetric): Coop vs. Defect	1.15	0.14	8.23	<.001
Non-correspondence (Low): Coop vs. Defect	1.15	0.14	8.21	<.001
Non-correspondence (Medium Low): Coop vs. Defect	0.87	0.14	6.30	<.001
Non-correspondence (Medium High): Coop vs. Defect	0.57	0.14	4.23	<.001
Symmetry (Symmetric) * Endowment (Low): Coop vs. Defect	0.74	0.20	3.76	<.001

Note. Coefficients denote log odds. For non-correspondence, the “High” condition is always the reference category.

Table 3.

Results of multinomial mixed effects model for the trinary-choice games, with "Defection" as the reference level for the outcome variable.

Fixed Effects	$\hat{\beta}$	se	*t*-value	*p*
Intercept: Coop vs. Defect	0.46	0.21	2.26	.024
Intercept: Exit vs. Defect	1.90	0.18	10.62	<.001
Endowment (Low): Coop vs. Defect	-1.30	0.20	-6.55	<.001
Endowment (Low): Exit vs. Defect	-1.27	0.17	-7.55	<.001
Symmetry (Symmetric): Coop vs. Defect	0.78	0.23	3.37	<.001
Symmetry (Symmetric): Exit vs. Defect	0.84	0.20	4.21	<.001
Non-correspondence (Low): Coop vs. Defect	1.67	0.22	7.62	<.001
Non-correspondence (Low): Exit vs. Defect	0.46	0.18	2.52	.012
Non-correspondence (Medium Low): Coop vs. Defect	1.58	0.22	7.26	<.001
Non-correspondence (Medium Low): Exit vs. Defect	0.48	0.18	2.65	.008
Non-correspondence (Medium High): Coop vs. Defect	0.64	0.21	3.03	<.001
Non-correspondence (Medium High): Exit vs. Defect	0.03	0.17	0.16	.876
Symmetry (Symmetric) * Endowment (Low): Coop vs. Defect	1.00	0.31	3.19	.001
Symmetry (Symmetric) * Endowment (Low): Exit vs. Defect	0.81	0.27	3.04	.002

Note. Coefficients denote log odds. For non-correspondence, the "High" condition is always the reference category.

Table 4.

Earnings of cooperators, earnings of defectors, and aggregated earnings of the players.

Matrix #	Earnings of cooperators			Earnings of defectors			Aggregated earnings of the two players		
	Binary-choice	Trinary-choice	Δ	Binary-choice	Trinary-choice	Δ	Binary-choice	Trinary-choice	Δ
1. High endowment, symmetric, low non-correspondence	14.02	13.06	-0.96	16.02	14.16	-1.86	28.93	23.55	-5.38
2. Low endowment, symmetric, low non-correspondence	7.16	6.48	-0.68	8.16	7.02	-1.14	14.72	11.7	-3.02
3. High endowment, asymmetric, low non-correspondence	12.13	12.17	0.04	14.13	13.23	-0.90	19.18	16.92	-2.26
4. Low endowment, asymmetric, low non-correspondence	4.83	5.27	0.44	5.83	5.91	0.08	15.89	16.02	0.13
5. High endowment, symmetric, medium low non-correspondence	12.05	11.82	-0.23	16.05	13.75	-2.30	26.07	22.02	-4.05
6. Low endowment, symmetric, medium low non-correspondence	6.31	5.89	-0.42	8.31	6.99	-1.32	13.47	11.2	-2.27
7. High endowment, asymmetric, medium low non-correspondence	9.52	11.44	1.92	13.52	13.63	0.11	16.71	16.59	-0.12
8. Low endowment, asymmetric, medium low non-correspondence	4.00	4.93	0.93	6.00	6.22	0.22	15.00	15.86	0.86
9. High endowment, symmetric, medium high non-correspondence	10.32	10.52	0.20	16.32	13.17	-3.15	23.79	20.93	-2.86
10. Low endowment, symmetric, medium high non-correspondence	5.23	5.22	-0.01	8.23	6.77	-1.46	11.98	10.59	-1.39
11. High endowment, asymmetric, medium high non-correspondence	7.36	9.40	2.04	13.36	12.59	-0.77	15.31	15.45	0.14
12. Low endowment, asymmetric, medium high non-correspondence	3.28	3.98	0.70	6.28	5.92	-0.36	14.62	14.97	0.35
13. High endowment, symmetric, high non-correspondence	7.76	9.36	1.60	15.76	12.64	-3.12	21.18	20.24	-0.94
14. Low endowment, symmetric, high non-correspondence	3.76	4.37	0.61	7.76	6.14	-1.62	10.50	10.07	-0.43
15. High endowment, asymmetric, high non-correspondence	5.87	8.90	3.03	13.87	12.86	-1.01	14.93	15.22	0.29
16. Low endowment, asymmetric, high non-correspondence	2.37	3.64	1.27	6.37	5.91	-0.46	14.37	14.87	0.50
Total (Matrix 1-16)	**115.97**	**126.45**	**10.48**	**175.97**	**156.91**	**-19.06**	**276.65**	**256.20**	**-20.45**

Note. A detailed description of these calculations is included in Appendix B. The earnings of participants who selected the exit option are fixed (see Appendix A), and thus independent of the other player's decision.

Figure 1. Akaike's Information Criterion (AIC) values for the different cluster solutions. Smaller AIC values indicate a better fit.

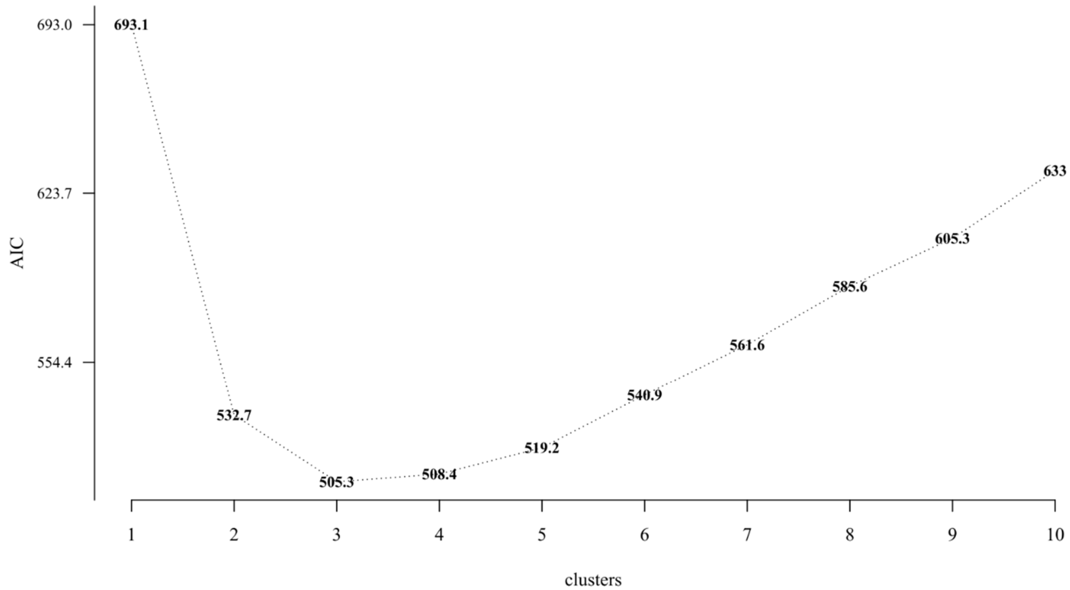

Figure 2. Main effect of endowment size on the predicted probabilities for participants' decisions in (A) the binary-choice games and (B) the trinary-choice games.

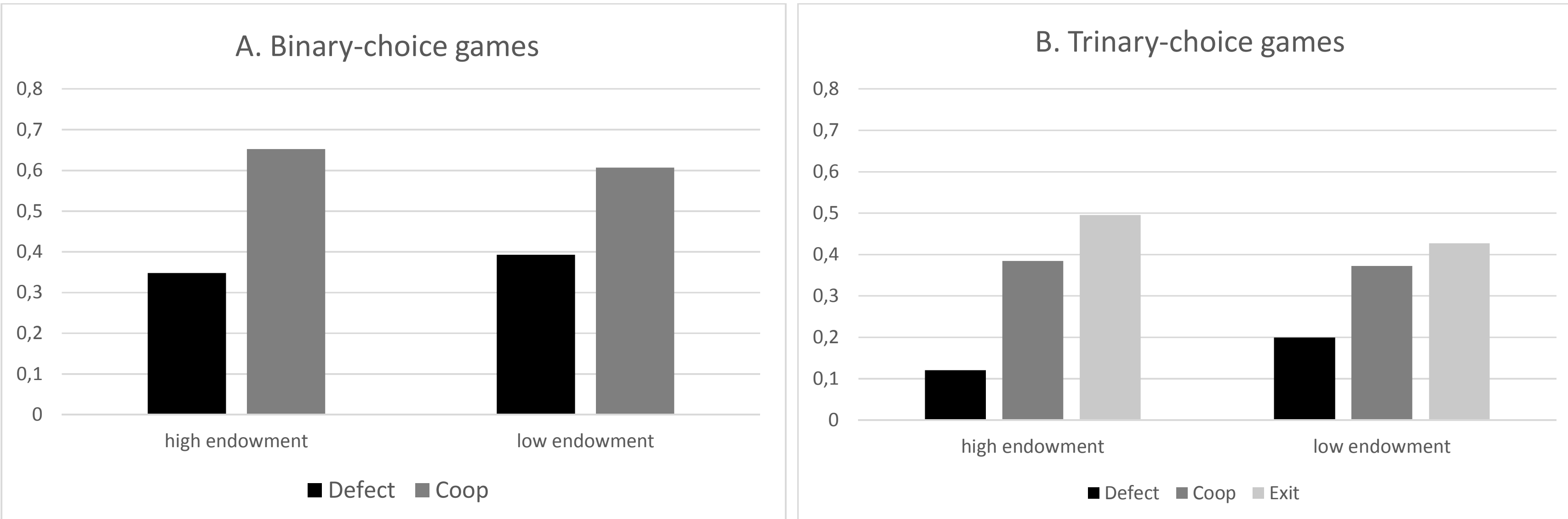

Figure 3. Main effect of (a)symmetry on the predicted probabilities for participants' decisions in (A) the binary-choice games and (B) the trinary-choice games.

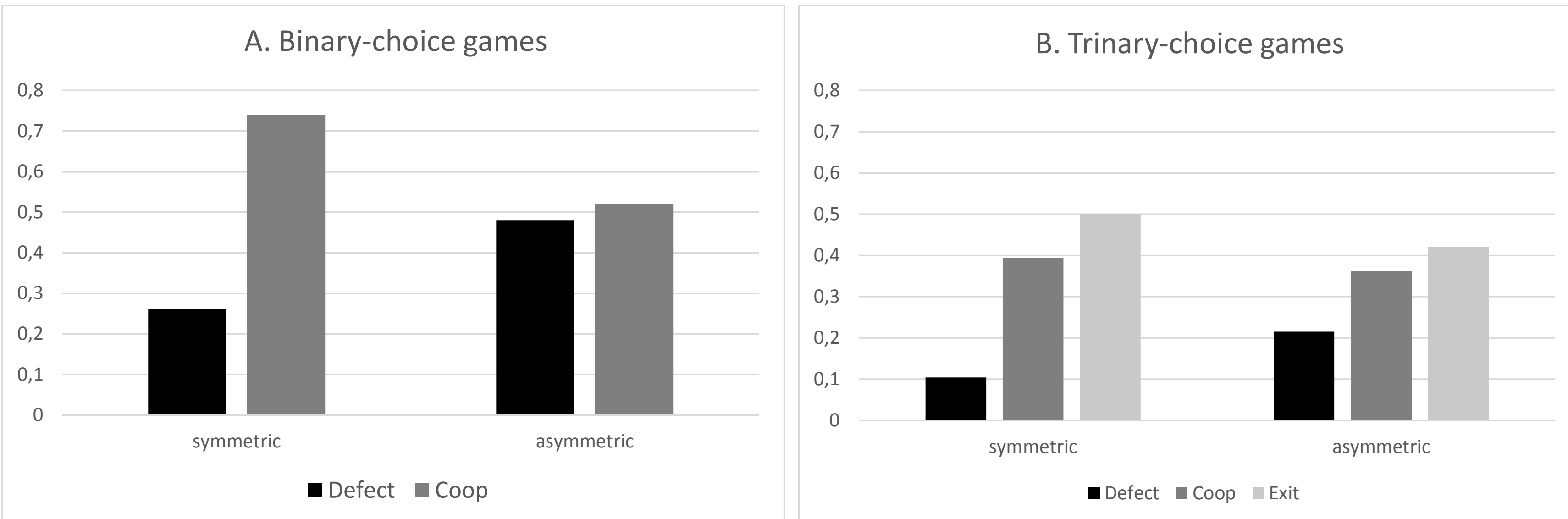

Figure 4. Interaction effect between endowment size and (a)symmetry on the predicted probabilities for participants' decisions in (A) the binary-choice games and (B) the trinary-choice games.

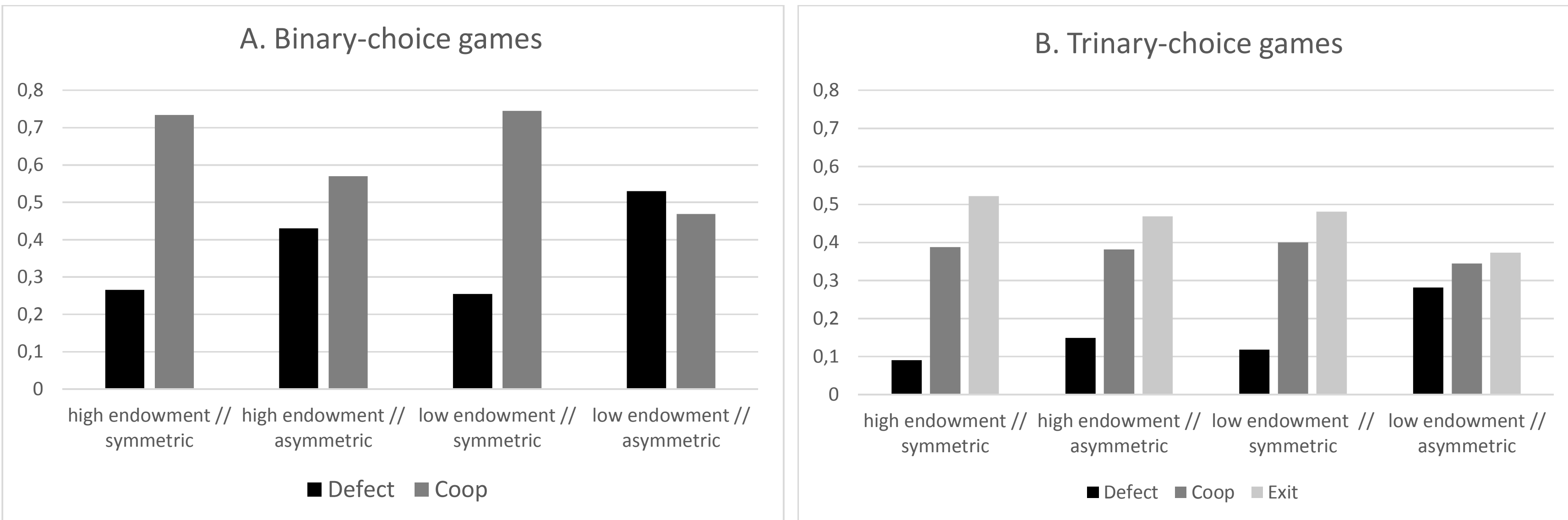

Figure 5. Main effect of non-correspondence on the predicted probabilities for participants' decisions in (A) the binary-choice games and (B) the trinary-choice games.