

In press: *Acta Psychologica*, © 2018, Elsevier. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors permission. The final article will be available, upon publication, via: <https://doi.org/10.1016/j.actpsy.2018.04.014> CC BY-NC-ND 4.0

Persistent Cooperation and Gender Differences in Repeated Prisoner's Dilemma Games: Some Things Never Change

Andrew M. Colman, Briony D. Pulford, and Eva M. Krockow

Department of Neuroscience, Psychology and Behaviour,
University of Leicester, Leicester LE1 7RH, UK

Abstract

In the finite-horizon repeated Prisoner's Dilemma, a compelling backward induction argument shows that rational players will defect in every round, following the uniquely optimal Nash equilibrium path. It is frequently asserted that cooperation gradually declines when a Prisoner's Dilemma is repeated multiple times by the same players, but the evidence for this is unconvincing, and a classic experiment by Rapoport and Chammah in the 1960s reported that cooperation eventually recovers if the game is repeated hundreds of times. They also reported that men paired with men cooperate almost twice as frequently as women paired with women. Our conceptual replication with Prisoner's Dilemmas repeated over 300 rounds with no breaks, using more advanced, computerized methodology, revealed no decline in cooperation, apart from endgame effects in the last few rounds, and replicated the substantial gender difference, confirming, in the UK, a puzzling finding first reported in the US in the 1960s.

PsycINFO Classification codes: 2340; 2970; 3020

Keywords: Cooperation; Endgame effect; Gender difference; Prisoner's Dilemma; Social dilemma

Email Andrew M. Colman (corresponding author): amc@le.ac.uk

Email Briony D. Pulford: bdp5@le.ac.uk

Email Eva M. Krockow: emk12@le.ac.uk

1. Introduction

The archetypal social dilemma is the two-player Prisoner's Dilemma (PD), a game that has been subjected to much experimental investigation in the history of experimental games and behavioral game theory (Rapoport, Seale, & Colman, 2015; Roth, 1995). Among the many reasons for its enduring popularity is the fact that it provides a conceptual structure within which phenomena such as cooperation and competition, trust and trustworthiness, altruism and spite, threats, promises, commitments, and collective rationality can be formalized and investigated rigorously, on the basis of behavioral measures rather than mere questionnaire responses (Pruitt & Kimmel, 1977; Rapoport & Chammah, 1965a); but what attracts researchers to it more than anything else is its paradoxical character and the challenge of explaining why players cooperate. In single-play (one-shot) and finite-horizon repeated PD, there are compelling arguments, explained in the Subsection 1.1 below, why rational players should never cooperate.

The research reported in this article focuses on cooperation in the finite-horizon repeated PD, in which the game is repeated over a finite number of rounds by the same players, who know in advance how many rounds will be played. Many experiments have addressed this issue, but the vast majority used only short sequences of repetitions. An exhaustive meta-analysis of experiments on trust and cooperation in both two-player and multi-player social dilemmas (Balliet & Van Lange, 2013) found 212 experiments, 132 using one-shot interactions and the rest only small numbers of repetitions ($M = 6.07$, $SD = 13.54$); hardly any used more than 50 repetitions (see also Balliet, Mulder, & Van Lange 2011; Embrey, Fréchette, & Yuksel, 2018).

The most ambitious experiment with long sequences or repetitions (Rapoport & Chammah, 1965a) involved 140 experimental subjects playing 300 rounds, and seven different PDs. The researchers reported an initial decline in cooperation followed by a recovery after many repetitions: “The most typical feature of the time course of a Prisoner’s Dilemma protocol is the initial decline in cooperation, followed eventually [after 30–60 rounds] by a recovery” (p. 200). However, there are some aspects of the experiment that make this conclusion difficult to interpret. The instructions given to the subjects began: “You will be playing a game,” probably priming an initially competitive mental set, because the objective in virtually all familiar indoor and outdoor games is to beat the opponent. The experiment included incentive payments, but the level of remuneration was 1/10 of a penny (US cent) per payoff point, derisory even in the 1960s. Above all, the “time courses”—the claimed declines and recoveries in the relative frequency of cooperative choices—were averaged over seven PDs with different payoffs and presented only graphically as moving averages. In Figure 7 (p. 90) and Figure 17 (p. 97), showing results for the relevant “pure matrix” treatment conditions, it is far from obvious that the reported initial decline is statistically significant, and no evidence is provided to back this up, because appropriate statistical techniques for analyzing time series had not yet been developed when the experiment was conducted.

The same classic study was also the first to report a large and unexpected gender difference, with male/male pairs cooperating almost twice as frequently as female/female pairs (Rapoport & Chammah, 1965b). Such large gender differences are seldom reported in psychology, and a natural expectation, based on traditional sex roles and socialization, would be of more cooperation in female/female than male/male pairs. The experiment reported below was designed as a conceptual replication to check Rapoport and Chammah’s findings using more rigorous experimental and data-analytic techniques and also to provide some evidence on the cross-cultural generalizability and temporal stability of these findings.

1.1. Theoretical considerations

Figure 1(a) shows the payoff matrix of a PD with payoff values originally introduced by Scodel, Minas, Ratoosh, and Lipetz (1959), popularized by Axelrod (1980a, 1980b, 1984), and nowadays frequently described as “conventional” (Press & Dyson, 2012). The original symmetric version used by Tucker (1950/2001) when he named the newly discovered game in 1950 is shown in Figure 1(c). Player I chooses between the rows marked C (cooperate) and D (defect), Player II independently chooses between columns C and D, and the cell in which the pair of strategy choices intersect is the outcome of the game, with the payoffs to Player I and Player II listed in that order by convention.

In the conventional version (Figure 1a), if the game is played just once, then both players do better if both cooperate (each receiving 3 units) than if both defect (each receiving 1 unit). Nevertheless, rational players are bound to defect, because D is a *dominant strategy* for both

players, yielding a higher payoff than the C strategy whether the co-player chooses C or D, and D is therefore an unconditionally best strategy. The (D, D) outcome, in which both players choose their optimal D strategies, is the unique *Nash equilibrium* of this game—the only outcome in which each player’s strategy is a best reply to the co-player’s, in the sense that no other strategy yields as high a payoff against the co-player’s chosen strategy. For example, the (C, D) outcome is out of equilibrium: Player I’s choice of C is not a best reply to Player II’s D, because Player I could have received a better payoff by choosing D, given Player II’s choice of D. Only in the (D, D) outcome are both players strategies best replies and hence in Nash equilibrium.

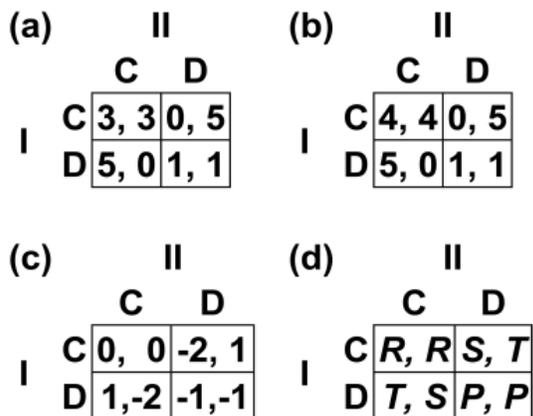


Figure 1. Prisoner’s Dilemma games with different payoff and index of cooperation (K) values. (a) Conventional version with $K = 2/5$ or 0.40; (b) Mild version with a higher index of cooperation, $K = 3/5$ or 0.60; (c) Original symmetric version from Tucker (1950/2001) with $K = 1/3$ or 0.33 approximately; (d) Generalized payoff matrix for any symmetric 2×2 game. The index of cooperation $K = (R - P)/(T - S)$.

In a repeated PD with no finite horizon or end-point known in advance, there are reasons to cooperate in spite of the dominance of the D strategy in the one-shot version, because rounds that have yet to be played cast a “shadow of the future” over earlier rounds. If Player I defects in Round t , then Player II may retaliate with defection in Round $t + 1$ or later, reducing Player I’s payoff. But in a repeated PD with a finite horizon—one in which a finite number of rounds are to be played and the players know this number—rational players will defect in every round. This is persuasively proved by the following argument (Luce & Raiffa, 1957, pp. 97–102; Sobel, 1993). Suppose the players know that there are to be exactly 100 rounds. In Round 100, there is no reason to cooperate, because there are no rounds to follow and therefore no possibility of retaliation; therefore, both players will defect in Round 100, because the D strategy is dominant, and both therefore do better by choosing D than C irrespective of what the co-player chooses. In Round 99, both players know that the outcome of Round 100 is predetermined, for the reason just given, therefore there is no reason to cooperate in Round 99, and players will choose their dominant strategies. This argument unfolds backwards in the same way, mandating defection in every round, including the first. Joint defection in every round is the only Nash equilibrium of the finite-horizon repeated PD, and it is proved by the argument above, called *backward induction*. However, the conclusion relies on full common knowledge of rationality. In a highly cited article, Kreps, Milgrom, Roberts, and Wilson (1982) showed that if both players are strictly rational payoff maximizers, but at least one believes that there is even a tiny probability that the

other is irrational, then rational cooperation can occur until close to the final round (for slightly different approaches, see Ambrus & Pathak, 2011; Dijkstra & Van Assen, 2017).

1.2. *Experimental evidence*

The first experimental study of the finite-horizon repeated PD was performed by Dresher and Flood in January 1950 and reported in a RAND research memorandum RM-789 in 1952, subsequently revised and condensed for publication by Flood (1958). Two research subjects, who were friends and had a knowledge of game theory, played exactly 100 incentivized rounds of an asymmetric PD in which $T > R > P > S$ for each player considered separately (see Figure 1d). The relative frequency of C choices was 73%, and “there was a decided tendency to start with [(D, D)] and then to shift to [(C, C)] rather consistently after about thirty trials” (pp. 14–15), except for the very last round, in which both players defected.

In an influential monograph on game theory, Luce and Raiffa (1957) suggested that repeated PDs should evolve toward joint cooperation in that way: “We feel that in most cases an unarticulated collusion between the players will develop. . . . This arises from the knowledge that the situation will be repeated and that reprisals are possible” (p. 101). This prediction appeared to be comprehensively refuted when the first full-scale, incentivized experiment was published two years later (Scodel, Minas, Ratoosh, & Lipetz, 1959). Introducing for the first time the conventional version of the game shown in Figure 1(a), their 22 player pairs (all men) completed 50 rounds of the game, and only two pairs showed evidence of collusion or increase in joint cooperation. Overall, significantly more D choices and joint defection outcomes (DD *lock-ins*) were observed in the last 25 rounds than in the first. Minas, Scodel, Marlowe, and Rawson (1960) and others replicated this effect, and an early review of published PD experiments concluded: “In general, the percentage of cooperative responses . . . tends to decrease over a series of trials” (Gallo & McClintock, 1965, p. 74). This finding has been replicated in more recent research; for example, Cooper, DeJong, Forsythe, and Ross (1996) reported: “Cooperation rates are positive and generally declining over time in the [finite-horizon repeated PD]” (p. 200). A simple learning model predicts just such a decline (Bornstein, Erev, & Goren, 1994), but the experimental studies that have shown a decline have not used long sequences of repetitions (e.g., Cooper et al. used only 10), and the declines reported in empirical studies may have been mere endgame effects as cooperation tends to disappear in the last few rounds.

The classic experiment by Rapoport and Chammah (1965a, 1965b) involved 300 rounds, and these researchers reported an initial decline followed by a slow recovery of cooperation. However, other well-controlled early experiments with long, finite sequences (e.g., Lave, 1962) reported no decline until the last few rounds, when an endgame effect was observed as cooperation declined sharply. More recent evidence (Normann & Wallace, 2012; Selten & Stoeker, 1986) has confirmed the late endgame effect. Taken together, the evidence suggests that the decline reported by some investigators may either have been illusory or may be fully explained by endgame effects in experiments using short sequences of repetitions.

Regarding the unexpected gender difference discovered by Rapoport and Chammah (1965b), a meta-analytic review of gender differences in social dilemmas (Balliet, Li, Macfarlan, & Van Vugt, 2011) corroborated this surprising finding and found that it increases in long sequences of repetitions: in 144 studies of PDs, “as people interact during repeated social dilemmas, over time men become increasingly more cooperative than women” (p. 888). Balliet et al. also reviewed results from public goods games, resource dilemmas, and N -player PDs and found gender differences in mixed male/female pairs as well as male/male and female/female

pairs. However, in the PD itself, Rapoport and Chammah (1965b) noted that “when men and women play with partners of the opposite sex, there is little or no discernible difference between them” (p. 837). There is a very high correlation (over .95) between the relative frequency of cooperative choices within pairs, especially in long sequences of rounds, because repeated defection by one player in the PD effectively forces the co-player to defect in order to avoid the sucker’s payoff. In Rapoport and Chammah’s examination of the PD, the gender difference was observable only in comparisons of male/male and female/female pairs, but was present in all seven PD games used.

Although the gender difference seems better corroborated than the time course of cooperation reported by Rapoport and Chammah (1965a), most of the evidence is quite old and is based mainly on North American studies, although Pansini, Shi, and Wang (2016) recently corroborated the gender difference strongly in southwest China. Furthermore, Rapoport and Chammah’s subjects knew the genders of their co-players, making it impossible to infer whether women cooperated less frequently in that study because they were women or because their co-players were women. Molina, Giménez-Nadal, Cuesta, Gracia-Lazaro, Moreno, and Sanchez (2013) reported results of an experiment in Spain in which male high-school students cooperated significantly *less* than female high-school students in 51 rounds of a multiplayer Prisoner’s Dilemma, apparently reversing Rapoport and Chammah’s finding, but an examination of the experimental game reveals that it had $T > R > P = S$ rather than $T > R > P > S$, (Figure 1d) and was thus not a Prisoner’s Dilemma with strongly dominant D strategies intersecting in a unique Nash equilibrium, but rather a “Goldenballs” game with three weak pure-strategy Nash equilibria. It seems worth checking whether the gender difference can be replicated in a present-day UK sample. Regarding the time course of cooperation, there is clearly insufficient evidence in the literature relating to long sequences of repetitions. Most early experiments have been characterized by few repetitions (50 or below), financial incentives that were usually either derisory or non-existent, absence of computerized methodologies enabling tight control of extraneous variables, and primitive data-analytic techniques. Later experiments, performed mostly by behavioral economists, focused not on finite-horizon PD but primarily on the *random termination* technique and usually examined effects over fewer than 25 iterations.

Turning to studies using the random termination technique, Dal Bó and Fréchette (2018) have comprehensively reviewed studies of “infinitely” repeated PDs, using a technique introduced by Roth and Murnighan (1978). In random termination studies, players never know whether any round is the last but always know the probability that a further round will be played. Studies using this technique have reported that the greater the probability of further rounds, the greater the cooperation, as should be expected. Almost half the results reviewed (18 out of 38 treatment conditions reported in 15 articles) used a probability of .75, implying that, although in theory the sequence could indeed continue forever, the probability of more than 25 rounds is less than one-tenth of 1%; it is therefore clear that hardly any long series of repetitions would have been played. Furthermore, from a psychological point of view, random termination is unlikely to be interpreted by experimental subjects as infinite repetition, because they are booked in advance for testing sessions and almost invariably know that the session must end within an allotted time slot. Normann and Wallace (2012) provided strong experimental evidence that random termination does not result in significantly more cooperation than finite-horizon games.

None of the early or later published experiments on repeated PD have included sophisticated data analysis of time courses. In most cases, changes in relative frequencies of cooperation over rounds were either (as in Rapoport & Chammah, 1965a) identified by

inspection, without any evidence of significance, or (as in Scodel, Minas, Ratoosh, & Lipetz, 1959) inferred by dividing rounds into two or more trial blocks and comparing the means either visually or with t tests or ANOVA. However, variables measured repeatedly on successive occasions are not stochastically independent, have correlated residuals or error terms, and typically display variances that do not remain stable across the series. For these reasons, they are not suited to conventional statistical procedures that assume independent and identically distributed data sets and independent residuals. The appropriate analytical methods are provided by time series analysis, specifically designed to yield estimates of the nature and strength of effects such as that of earlier decisions on later decisions and of the form of the overall trend over rounds (Chatfield, 2003; Yaffee & McGee, 2000; Yanovitzky & VanLear, 2008).

2. Present investigation

To provide an improved conceptual replication of the study reported by Rapoport and Chammah (1965a, 1965b), and thereby to establish the robustness across time, space, and experimental methodology of the time course and gender differences reported in that classic study, we carried out a computer-controlled experiment over 300 rounds of finite-horizon PDs with players in fixed pairs, and we analyzed the relative frequency of cooperative choices using time series analysis. This also afforded an independent check of the endgame effect identified by Lave (1962), Selten and Stoecker (1986), and Normann and Wallace (2012). We investigated two experimental PDs with different payoffs and index of cooperation values (Figs. 1a and 1b) to check whether the results would be replicated in a mild version with a higher index of cooperation than the conventional version. To determine whether the unexpected and remarkably large gender difference first identified by Rapoport and Chammah (1965b) could be replicated in a contemporary UK population, with players unaware of the genders of their co-players, we included both male and female experimental subjects and anonymous pairing. Lastly, as an aid to interpretation of results, we measured response times for strategy choices to provide an objective index of the amount of thought that players devoted to different stages of the game.

2.1. Method

2.1.1. Subjects

The sample comprised 150 subjects (60 males and 90 females) with a mean age of 22.80 years ($SD = 7.33$). For a medium effect size, power = 0.8, we required $N \geq 128$. Recruitment was through notices in the university's weekly electronic newsletter. In game theory, payoffs are *utilities* representing the players' true preferences, determined or revealed by their actual choices, although in most experimental games the payoffs are simply monetary values to be gained or lost. Incentives are necessary in experimental games, although they do not have to be large (Pulford, Colman, & Loomes, 2018). Our subjects received the cash equivalent of the payoffs from two randomly selected game outcomes during the testing session. Incentive earnings per person ranged from £0 to £10 (\$13.33), with an average of £5.97 (\$7.96). This is called the *within-subjects random lottery incentive system*, and the technique successfully eliminates problems associated with other remuneration schemes (Lee, 2008), avoids giving subjects the idea that any particular decision has only small material incentives, and has been shown empirically to elicit true preferences (Cubitt, Starmer, & Sugden, 1998; Starmer & Sugden, 1991). In addition to the incentive earnings, every participant received a £3.00 show-up fee to remunerate them for their time.

2.1.2. Design

The experimental design was a 2 (Game Treatment) \times 3 (Gender Composition) factorial design. Subjects were randomly assigned to player pairs, and the pairs were randomly assigned to two treatment conditions with different PDs. The second between-subjects variable was the gender composition of player pairs, with three levels: all-female (33 pairs), mixed-gender (24 pairs), and all-male (18 pairs). Each subject was unaware who the co-player was and therefore also of the gender of the co-player. The principal dependent variable was the total number of cooperative (C) choices per player pair over 300 rounds of the game. Additionally, response times (in ms) were recorded to assess how long subjects took to make their decisions. No variables or participants were discarded.

2.1.3. Materials

We used two experimental games: the conventional version (Press & Dyson, 2012) with index of cooperation $K = 0.40$ (Figure 1a), and the mild version with a higher index of cooperation $K = 0.60$ (Figure 1b). PDs with high K elicit more cooperation than those with low K (Murphy & Ackerman, 2015; Rapoport, 1967). Testing sessions were conducted in a computer laboratory with desks separated by partitions, and each subject was paired randomly with another anywhere in the room. Subjects received instruction handouts containing detailed explanations of the games (see [online instruction materials](#): Colman, Pulford, & Krockow, 2017). They interacted through computer terminals controlled by z-Tree software (Fischbacher, 2007) and were provided with feedback about the choices and payoffs of both players after each round. They were informed at the start that exactly 300 rounds would be played, and in each round they were reminded of the current round number.

2.2. Procedure

Testing sessions typically involved four or five player pairs and lasted between 40 and 80 minutes. After completing consent forms, subjects read through the instructions in their own time. They were encouraged to ask questions, and these were answered in private. Once the computer program was started, subjects were assigned player roles (*Red* or *Blue*, to avoid any implication of precedence/subservience from labels such as *Player 1* and *Player 2*) and randomly matched with players of the opposite color. Player pairs were then randomly assigned to treatment conditions (conventional or mild PD), and each subject was unaware of the gender of the co-player for the testing session. Subjects remained within their roles (Red or Blue) and player pairs for the whole experiment, each player pair completing 300 rounds of the same PD (either conventional or mild), and made their choices simultaneously on keyboards by pressing J or K for a C or D choice respectively. Words such as “game” and “player,” which tend to prime competitive behavior, were avoided. On completion, the monitors displayed the subject’s final earnings, including show-up fee and earnings from two randomly selected games, and subjects were paid in cash before leaving the laboratory.

3. Results

3.1. Game and gender effects

All raw experimental data are accessible online (Colman, Pulford, & Krockow, 2017). We first compared mean cooperation rates across treatment conditions. For the statistical analysis in this subsection, the unit of analysis is the player pair, because decisions within pairs are not stochastically independent.

The mean number of cooperative choices per player pair out of a possible maximum of 300 was 227.10, indicating that 75.70% of choices were cooperative, 95% CI [208.59, 245.61]. The mean for the conventional version of the game was 221.51, 95% CI [194.47, 248.55], and the mean for the mild version was 232.69, 95% CI [206.23, 259.15], indicating slightly more cooperation in the mild version, although this difference is not significant: $F(1, 68) = 0.43, p = .52$, partial $\eta^2 = .01$. Twenty-one of the individual players cooperated in all 300 rounds, and 17 of these were women; seven cooperated in 299 rounds but defected in the last, and six of these were women. This suggests that some of the most extremely cooperative individuals were women, in contradistinction to the female/female versus male/male player pair differences presented in the following paragraph; but these subsamples are too small to infer that this is a general phenomenon.

Comparing player pairs by gender composition, the means were 206.56, 95% CI [170.96, 242.17] in female/female pairs, 238.90, 95% CI [214.57, 263.22] in mixed-gender pairs, and 250.32, 95% CI [224.48, 276.17] in male/male pairs. Because the gender difference tends to be suppressed in mixed-gender pairs (Rapoport & Chammah, 1965b), we compared female/female with male/male pairs, and found that the female/female pairs were significantly less cooperative, $t(48) = 2.05, p = .046$ (two-tailed), $d = 0.55$ —a medium to large effect size. A planned comparison of female/female pairs with mixed-gender and male/male pairs combined confirmed that the female/female pairs were significantly less cooperative: $F(1, 70) = 3.98, p = .05$, partial $\eta^2 = .05$. No significant interaction between game and gender composition conditions was found.

Table 1

The Proportions of Players Choosing Different Options, Split by Gender Composition

	M/M		F/M		F/F	
	<i>Conv</i>	<i>Mild</i>	<i>Conv</i>	<i>Mild</i>	<i>Conv</i>	<i>Mild</i>
C	.77	.87	.77	.82	.69	.69
C(1)	.72	.72	.63	.79	.82	.72
C(2)	.56	.61	.63	.71	.74	.69
CC	.66	.83	.68	.75	.61	.62
CD/DC	.22	.08	.18	.14	.16	.13
DD	.12	.09	.14	.11	.23	.25
C(last 3)	.42	.48	.56	.61	.49	.58
<i>LCC</i>	.25	.44	.42	.50	.47	.50
<i>LDD</i>	.00	.00	.00	.00	.12	.06

Emulating the analysis by Rapoport and Chammah (1965b), Table 1 shows the proportion of each group that cooperated (C), mutually cooperated (CC), defected (DD), or where only one of the pair cooperated (CD/DC) over 300 rounds. It also shows the proportion that cooperated at the start of the series, Round 1 [C(1)] and Round 2 [C(2)], and at the very end in the last three rounds [C(last 3)]. Also following Rapoport and Chammah's analysis, Table 1 shows the proportion of pairs that locked into mutual cooperation (*LCC*) or mutual defection (*LDD*) in at least 23 of the final 25 rounds. Our data, like those of Rapoport and Chammah, showed virtually no differences between the genders in the mixed pairs, so we combined them into F/M for pairs with one man and one woman. It's not surprising that females with male partners, and males with female partners behaved the same when in a pair, because the strategic structure of the Prisoner's Dilemma causes the behavior of paired players to become very similar

over repeated rounds. Rapoport and Chammah explained that “paired players are pushed in the direction of more cooperation or in the direction of less cooperation, and whichever is the tendency, the paired players tend to be pushed in the same direction and so to become more like each other in the course of the session” (p. 836). In our experiment, as in Rapoport and Chammah’s, the correlation between the relative frequency of cooperative strategy choices within pairs was over .95, and this makes it impossible for sharp gender differences to emerge in mixed-gender pairs.

Cooperation is high in Round 1 and slightly less so in Round 2. Very few pairs were locked into mutual defection at the end of the session and many were mutually cooperating for much of the last 25 rounds. There was no significant impact of game or gender composition on the level of C choices in the last three rounds. Analysis of variance of the pairs’ choices over 300 rounds showed that the average number of CC, CD/DC, DD choices was not affected significantly by game treatment or gender composition and there was no interaction between them, but the number of CC choices was significantly higher than the other two types, $F(2, 138) = 72.22, p < .001, \text{partial } \eta^2 = .51$.

Table 2

The Probability of Players Choosing Different Options Given the Prior Round Outcome, Split by Gender

	$P(C DD)$		$P(C DC)$		$P(C CC)$		$P(C CD)$	
	w_i		z_i		x_i		y_i	
	<i>Conv</i>	<i>Mild</i>	<i>Conv</i>	<i>Mild</i>	<i>Conv</i>	<i>Mild</i>	<i>Conv</i>	<i>Mild</i>
Males	.25	.15	.50	.49	.94	.98	.55	.58
Females	.15	.15	.43	.34	.96	.97	.49	.43

In Table 2 we show the probability of players cooperating, given the outcome of the previous round. This is calculated over 299 rounds, because there is no round before Round 1. The probability that the players cooperated after they defected in the previous round is shown separately, depending on whether the co-player defected in that round [$P(C|DD)$] or cooperated [$P(C|DC)$]. Similarly the probability that the players cooperated after mutual cooperation in the previous round [$P(C|CC)$] is distinguished from when the player cooperated in the previous round but the co-player defected [$P(C|CD)$]. In Rapoport and Chammah’s (1965a) terminology these equate to w_i, z_i, x_i, y_i , respectively. Analysis of variance by player was not possible due to missing data (e.g., 21 participants never defected). Table 2 shows that players almost always followed a joint CC outcome with another C move in the next round. If both players had defected in the previous round, then a player was unlikely to choose C in the next round, with only about .15 to .25 probability of a C move. The probabilities of cooperation immediately following an unreciprocated defection choice (in the range .34 to .50) and immediately following an unreciprocated cooperative choice (.43 to .58) were both slightly higher. The probabilities of $P(C|DC)$ and $P(C|CD)$ seem to show the largest gender difference, especially in the mild condition where women show low levels of playing C when the previous round showed that the pair were not in sync.

We wanted to determine whether the length of time players spent deciding on their current choice was influenced by the outcome of the previous round, but statistical analysis was not straightforward, because 21 players cooperated in all 300 rounds. This resulted in missing response times after two types of outcome (DD, DC) for those unconditional cooperators, and

other players also had missing data for one or more of the four outcomes. To ameliorate this we combined data from the three outcomes that were *not* CC, because their means were similar, and we compared this combined score to the CC outcome. This resulted in complete data for 122 participants that we entered into a repeated measures ANOVA. The analysis revealed that the outcome of the previous round interacted with the game condition, $F(1, 118) = 8.41, p = .004$, partial $\eta^2 = .07$, and the player's gender had no influence. Players responded very quickly after a CC outcome in the previous round ($M = 0.35s, SD = 0.36s$ in the conventional version, and $M = 0.31s, SD = 0.28s$ in the mild version). All other outcomes in the previous round (DD, CD, and DC combined) resulted in taking significantly longer to respond in the next round, and this difference was greater in the mild version ($M = 1.23s, SD = 1.67s$) than in the conventional version ($M = 0.60s, SD = 0.52s$).

3.2. Time series analysis

Figure 2 shows sequence plots of the mean proportions of cooperative choices per player per round for female/female, male/male, and mixed-gender player pairs in the conventional version and the mild version. Visual inspection suggests no obvious decline in cooperation in early rounds or overall, although a large endgame effect, shown by a sharp decline in cooperation in the last few (2–4) rounds, is evident in all six time series. To identify and characterize long-term changes in cooperation over the full 300-round sequences, we performed time series analysis and selected the best-fitting model for each time series.

We begin with the conventional version of the game. For the data from female/female player pairs, ARIMA and exponential smoothing models were considered, and the best fitting model, according to the stationary R^2 and Ljung–Box criteria, was an ARIMA(2,0,0) model, suggesting that this time series is strongly autocorrelated, cooperation in each round being influenced significantly by the cooperation in the two previous rounds, but without any significant long-term trend. The best R^2 index of goodness of fit for time series is the stationary R^2 statistic; in this case stationary $R^2 = .39$, and the statistic Ljung–Box $Q(16) = 20.63, p = .19$, indicates that no significant temporal structure in the time series is unaccounted for by the ARIMA(2,0,0) model. For male/male player pairs, a simple Holt model—an exponential smoothing model in which the parameters are level and trend, unconstrained by each other's values—provided the best fit, indicating a linearly increasing trend in cooperation across the time series. For this time series, stationary $R^2 = .22$ and Ljung–Box $Q(17) = 17.82, p = .40$, confirming that no significant temporal structure is unexplained. For mixed-gender player pairs, a simple mean model provided the best fit, indicating no significant trend, with stationary $R^2 = .18$ and Ljung–Box $Q(17) = 25.13, p = .09$, confirming once again that no significant temporal structure is unaccounted for.

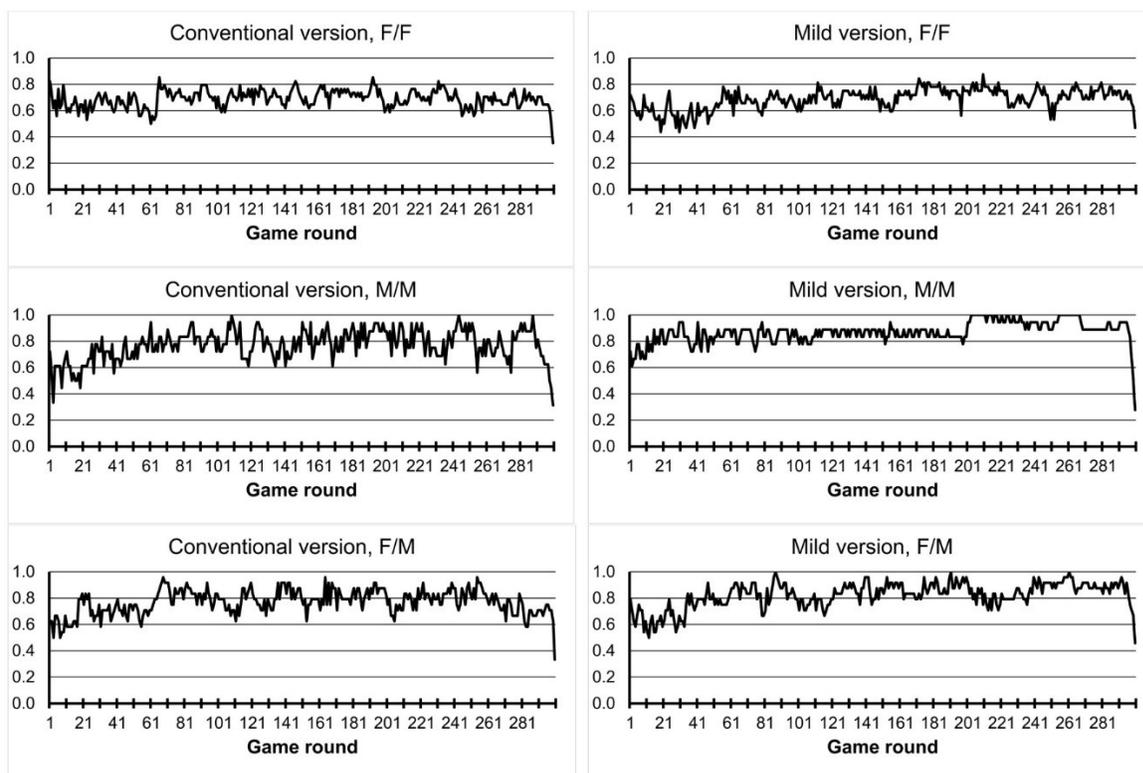


Figure 2. Sequence plots of mean proportions of cooperative choices per player per round for female/female (F/F), male/male (M/M), and mixed-gender (F/M) player pairs in a conventional version, Figure 1(a), and a mild version with a higher index of cooperation K , Figure 1(b). Cooperation in the round is coded as “1” and defection is coded as “0”.

Statistical power analysis revealed that time series with 250 or more observations were required to achieve satisfactory power $(1 - \beta) \geq .80$ at significance level $\alpha = .05$, using the generally preferred Ljung–Box Q statistic as a criterion of model fit. Our power analysis was based on large-scale computationally intensive simulations reported by McGee (2000), showing that the Ljung–Box test is able to detect *underdifferencing* in almost all cases with even fewer observations. This type of model misspecification includes a failure to model an overall trend, such as a decline in cooperation over repetitions—the primary focus of our interest in this research—when it is actually present in the data. Over 250 observations ensure that the Ljung–Box test almost invariably rejects the model fit in these cases, hence our time series, with 300 observations, have sufficient power to detect any trends.

Next, we examined data from the mild version of the game. For all gender pairings, simple mean models, indicating no significant long-term trend, provided the best fit. For female/female player pairs, stationary $R^2 = .21$, $Q(17) = 33.60$, $p = .009$; for male/male player pairs, stationary $R^2 = .04$, $Q(17) = 18.65$, $p = .35$; and for mixed-gender player pairs, stationary $R^2 = .09$, $Q = 11.04$, $p = .86$. Although these simple mean models do not explain much of the variance in cooperation, because stationary R^2 values are low, except in the female/female pairs, the Ljung–Box values suggest that they do not leave much unexplained temporal structure either.

We also performed time series analysis on decision times. Figure 3 displays the sequence plot of response times for all players in both games. For both games and all gender pairings, mean decision times were about 10 seconds in the first round, falling steeply and settling well

below 1 second after about 30 rounds, then rising in the last few (2–4) rounds, in line with the endgame effect in cooperation mentioned earlier. In all six Game \times Gender Composition conditions, the best fit was provided by a damped trend exponential smoothing model, applicable to a time series displaying a trend that gradually flattens out over time. For the aggregated series displayed in Figure 3, the three estimated parameters of the model were $\alpha = .697$, $t = 10.81$, $p < .001$; $\gamma = .999$, $t = 3.92$, $p < .001$; $\phi = .801$, $t = 18.18$, $p < .001$; α representing the smoothing parameter, γ the trend parameter, and ϕ the damping parameter.

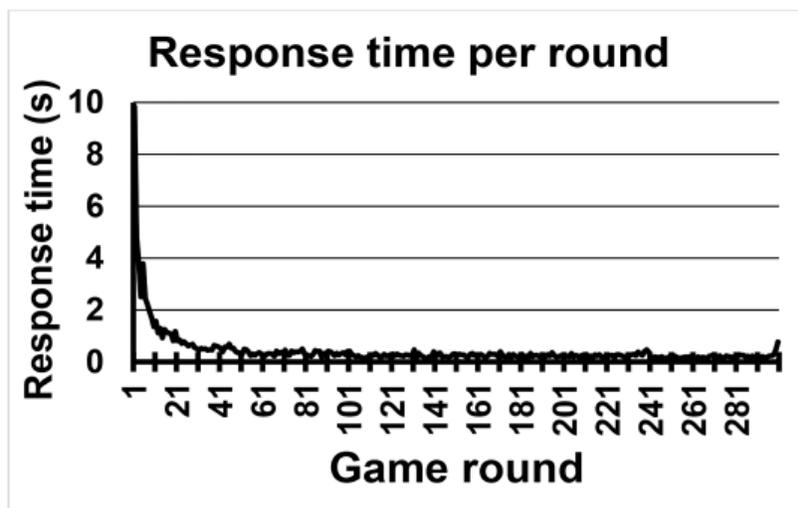


Figure 3. Sequence plot of response times in seconds over 300 rounds for all players in both games.

3.3. Endgame effects

We analyzed the cooperation rates in the last ten rounds to determine if endgame effects appeared earlier in the male/male pairs than the female/female or mixed-gender pairs. While cooperation significantly declined in the last 4 rounds, this was not significantly affected by the pair's gender, the player's gender, or the type of game played. We specifically examined the response times in the last ten rounds using ANOVA, and found that the response times rose significantly in the last 2 rounds (all $p < .01$), but that this was not differentially influenced by the pair's gender composition or the type of game played.

4. Discussion

Using a finite-horizon repeated Prisoner's Dilemma with conventional payoff values and also a mild version with a higher index of cooperation, we found no evidence of any decline in cooperation over 300 rounds in either game, although we used time series analysis powerful enough to detect even subtle trends. In the conventional game, there was evidence of a significant linear increase in cooperation in male/male pairs, though not in female/female or mixed-gender pairs. There was no significant increase in cooperation in the mild version, at least partly because of a ceiling effect arising from the very high levels of cooperation for male/male pairs in that game. We believe that the initial decline in cooperation reported by Rapoport and Chammah (1965a) on the basis of visual inspection of graphs, and the general decline reported by other investigators, may have been illusory, or artifacts caused by framing the task as a

competitive game and using derisory or non-existent incentives, or perhaps mere endgame effects when only small numbers of repetitions were investigated.

Across games, female/female pairs cooperated significantly less than male/male pairs, confirming that the gender difference discovered by Rapoport and Chammah (1965b) is still evident in a UK population and with a similar effect size ($d = .55$ in our study versus 0.57 in theirs). This suggests that researchers need to be attentive to the gender of players in experimental games, even when the players are ignorant of the gender of their co-players, as they were in the experiment reported here. We have replicated the gender difference in a country halfway round the world from the US, using a different and more rigorous experimental methodology, and above all in a different historical period, after a truly revolutionary change in gender roles among university students. In the US in the mid-1960s, before the rise of feminism, Ivy League universities such as Yale, Princeton, Columbia, and Dartmouth did not admit female students, and many other universities had Home Economics courses for women; at the University of Michigan, where Rapoport and Chammah performed their experiment, women had only recently been allowed to enter the Michigan Union building through the front door. In 1965, birth control was illegal for unmarried women in 26 states, women could not serve on federal juries in many states, and leading banks would not issue a credit card to a woman without her husband's signature. Despite a radical change in the status of women and in attitudes towards women, the gender difference in cooperation in repeated Prisoner Dilemma replicates unambiguously in the UK today.

Our gender difference was substantial and striking, and much larger than most gender differences in psychology. The effect is hard to explain, because it seems to contradict widely accepted assumptions about sex roles, according to which women are generally expected to be less competitive and more altruistic than men (Andreoni & Vesterlund, 2001; Niederle & Vesterlund, 2007). The most tempting explanation is that women are more risk-averse than men (Charness & Gneezy, 2012; Myers-Levy & Loken, 2015) and therefore tend to avoid the C strategy because it entails the risk of the worst (sucker's) payoff (Simpson, 2003). But this is contradicted by the finding that women also defect more frequently than men in the Chicken game, in which the D strategy risks the worst payoff (Capon-Dhersin & Etchart-Vincent, 2013; Rapoport & Chammah, 1969). A more persuasive explanation lies in the social structural distinction between agentic and communal characteristics (Eagly & Wood, 1999; Wood & Eagly, 2015). Typical family and economic roles of men and women can be described as resource provider and homemaker respectively, and this may cause women to acquire more communal patterns of behavior than men. Women may consequently be less achievement oriented than men, in terms of maximizing their personal payoffs, and more socially oriented, hence more concerned than men about relative rather than absolute payoffs. This may cause them to play "defensively" in a repeated PD by avoiding the possibility of earning less than their co-players (Hottes & Kahn, 1974). In both the PD and Chicken games, the strategy that avoids any possibility of relative loss is the D strategy. This explanation is therefore consistent with the experimental evidence, and also with the finding that women are more frequently than men motivated to avoid (relative) failure rather than to seek success (Horner, 1972; Ivers & Downes, 2012).

In Figure 2, an endgame effect is apparent in all treatment conditions, corroborating earlier findings of Lave (1962), Selten and Stoeker (1986), and Normann and Wallace (2012): men and women in both the conventional game and the mild version, although they cooperated throughout most of the game, showed sharply decreased cooperation in the last few rounds,

presumably because, with a vanishing shadow of the future, reasons for cooperating began to disappear. In line with this finding, response times, which in all treatment conditions were highest (around 10 seconds) in the first round and then decayed over the next 30 rounds to under one second, remained low but rose slightly in the last few rounds (typically by about 100 to 700 milliseconds). These findings suggest that players thought carefully before making their first strategy choices and then gradually speeded up as they became familiar with the problem, but that something caused them to pause and think again in the last few rounds. The late increase in response times is consistent with our interpretation that reasons for cooperating tend to evaporate as the end looms into view.

Our most important conclusion is that the frequently claimed decline in cooperation in repeated PDs appears to be a misconception, perhaps fueled by the assumptions of game theorists that there ought to be a steady convergence toward Nash equilibrium over rounds. Such convergence does indeed tend to occur when players have experience of multiple repeated PDs with different co-players (Embrey, Fréchette, & Yuksel, 2018, Selten & Stoecker, 1986); but within a single repeated game with a fixed co-player, there is no evidence that cooperation declines systematically until endgame effects are activated as the final round approaches. Some of the earliest experiments that used long sequences of repetitions (Flood, 1958; Lave, 1962) yielded results very similar to our own. In short sequences of repetitions, endgame effects may begin earlier and may then be misinterpreted as a general decline in cooperation. Also, quasi-endgame effects may occur as methodological artifacts. For example, Rapoport and Chammah (1965a) informed their participants that “After each series of 25 moves you will be asked to total your gains and losses” (p. 228), and this may have created quasi-endgame effects unintentionally at earlier points in their 300 rounds. The initial decline in cooperation, of around 5–6% in their first 25 rounds, shown in Figure 7 (p. 90), may be due to a quasi-endgame effect in rounds 23–25, which would have shown up in the “running averages” of 15 games, on the x -axis, from point 9 onwards (that averages rounds 9–23). A drop in cooperation in rounds 23–25 of around 20%, such as we found occurring as an endgame drop, would show up in the 15 round averages from rounds 9 to 25 as a drop of 5%, accounting for their initial downward slope in cooperation in Figure 7. Recovery from this endgame effect would take several rounds to occur because, as we found, the probability of C after DD is only .15 to .25. Only an examination of Rapoport and Chammah’s raw data, or an exact replication of their procedure, could determine whether our speculation is correct.

The persistence of cooperation that we observed applies to the conventional PD and also to a milder version that we tested; we would not wish to claim that no decline in cooperation would occur over repeated rounds of a more severe PD with a low cooperation index (see Figure 1). That is a question that can be answered only by further experimental research. Future researchers must be careful not to introduce quasi-endgame behavior unintentionally at earlier points in their testing sessions by having anticipated breaks in the sequence of rounds, which may generate declines in cooperation as mere methodological artifacts.

Acknowledgements

The research was supported by Grant RM43G0176 from the Leicester Judgment and Decision Making Endowment Fund.

References

- Ambrus, A., & Pathak, P. A. (2011). Cooperation over finite horizons: A theory and experiments. *Journal of Public Economics*, 95(7–8), 500–512. doi:10.1016/j.jpubeco.2010.11.016
- Andreoni, J., & Vesterlund, L. (2001). Which is the fair sex? Gender differences in altruism. *Quarterly Journal of Economics*, 116(1), 293–312. doi:10.1162/003355301556419
- Axelrod, R. (1980a). Effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, 24(1), 3–25. doi:10.1177/002200278002400101
- Axelrod, R. (1980b). More effective choice in the Prisoner's Dilemma. *Journal of Conflict Resolution*, 24(3), 379–403. doi:10.1177/002200278002400301.
- Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.
- Balliet, D., Li, N. P., Macfarlan, S. J., & Van Vugt, M. (2011). Sex differences in cooperation: A meta-analytic review of social dilemmas. *Psychological Bulletin*, 137(6), 881–909. doi:10.1037/a0025354
- Balliet, D., Mulder, L. B., & Van Lange, P. A. M. (2011). Reward, punishment, and cooperation: A meta-analysis. *Psychological Bulletin*, 137(4), 594–614. doi:10.1037/a0023489
- Balliet, D., & Van Lange, P. A. M. (2013). Trust, conflict, and cooperation: A meta-analysis. *Psychological Bulletin* 139(5), 1090–1112. doi:10.1037/a0030939
- Bornstein, G., Erev, I., & Goren, H. (1994). The effect of repeated play in the IPG and IPD games. *Journal of Conflict Resolution*, 38(4), 690–707. doi:10.1177/0022002794038004005
- Cabon-Dhersin, M.-L., & Etchart-Vincent, N. (2013). Cooperation: The power of a single word? Some experimental evidence on wording and gender effects in a game of chicken. *Bulletin of Economic Research* 65(1), 43–64. doi:10.1111/j.1467-8586.2012.00468.x
- Charness, G., & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, 83(1), 50–58. doi:10.1016/j.jebo.2011.06.007
- Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). New York: Chapman & Hall/CRC.
- Colman, A. M., Pulford, B. D., & Krockow, E. M. (2017). *Repeated Prisoner's Dilemma dataset*. Leicester Research Archive. <https://lra.le.ac.uk/handle/2381/40367> [dataset]
- Cooper, R., DeJong, D. V., Forsythe, R., & Ross, T. W. (1996). Cooperation without reputation: Experimental evidence from Prisoner's Dilemma games. *Games and Economic Behavior*, 12(2), 187–218. doi:10.1006/game.1996.0013
- Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics*, 1(2), 115–131. doi:10.1007/BF01669298
- Dal Bó, P., & Fréchet, G. R. (2018). On the determinants of cooperation in infinitely repeated games: A survey. *Journal of Economic Literature*, 56(1), 60–114. doi:10.1257/jel.20160980
- Dijkstra, J., & Van Assen, A. L. M. (2017). Explaining cooperation in the finitely repeated simultaneous and sequential prisoner's dilemma game under incomplete and complete information. *Journal of Mathematical Sociology*, 41(1), 1–25. doi:10.1080/0022250x.2016.1226301
- Eagly, A. H., & Wood, W. (1999). The origins of sex differences in human behavior: Evolved dispositions versus social roles. *American Psychologist*, 54, 408–423. doi:10.1037/0003-066X.54.6.408

- Embrey, M., Fréchette, G. R., & Yuksel, S. (2018). Cooperation in the finitely repeated Prisoner's Dilemma. *Quarterly Journal of Economics*, *133*(1), 509–551. doi:10.1093/qje/qjx033
- Fischbacher, U. (2007). z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. doi:10.1007/s10683-006-9159-4
- Flood, M. M. (1958). Some experimental games. *Management Science*, *5*(1), 5–26. doi:10.1287/mnsc.5.1.5
- Gallo, P. S., & McClintock, C. G. (1965). Cooperative and competitive behavior in mixed-motive games. *Journal of Conflict Resolution*, *9*(1) 68–78. doi:10.1177/002200276500900106
- Horner, M. S. (1972). Toward an understanding of achievement-related conflicts in women. *Journal of Social Issues*, *28*(2), 157–175. doi:10.1111/j.1540-4560.1972.tb00023.x
- Hottes, J. H., & Kahn, A. (1974). Sex differences in a mixed-motive conflict situation. *Journal of Personality*, *42*(2), 260–275. doi:10.1111/j.1467-6494.1974.tb00673.x
- Ivers, J. H., & Downes, P. (2012). A phenomenological reinterpretation of Horner's fear of success in terms of social class. *European Journal of Psychology of Education*, *27*(3), 369–388. doi:10.1007/s10212-011-0076-3
- Kreps, D. M., Milgrom, P., Roberts, J., Wilson, R. (1982). Rational cooperation in the finitely repeated prisoners' dilemma. *Journal of Economic Theory*, *27*(2), 245–252. doi:10.1016/0022-0531(82)90029-1
- Lave, L. B. (1962). An empirical approach to the Prisoner's Dilemma game. *Quarterly Journal of Economics*, *76*(3), 424–436. doi:10.2307/1879629
- Lee, J. (2008). The effect of the background risk in a simple chance improving decision model. *Journal of Risk and Uncertainty*, *36*(1), 19–41. doi:10.1007/s11166-007-9028-3
- Luce, R. D., & Raiffa, H. (1957). *Game and decisions: Introduction and critical survey*. New York: Wiley.
- McGee, M. (2000). Power analysis and sample size determination for well-known time series models. In R. A. Yaffee & M. McGee, *Introduction to time series analysis and forecasting: With applications of SAS and SPSS* (pp. 481–493). Cambridge, MA: Academic Press.
- Minas, J. S., Scodel, A., Marlowe, D., & Rawson, H. (1960). Some descriptive aspects of two-person non-zero-sum games. II. *Journal of Conflict Resolution*, *4*(2), 193–197. doi:10.1177/002200276000400204
- Molina, J. A., Giménez-Nadal, J. I., Cuesta, J. A., Gracia-Lazaro, C., Moreno, Y., & Sanchez, A. (2013). Gender differences in cooperation: Experimental evidence on high school students. *PLOS ONE*, *8*(12), e83700. doi:10.1371/journal.pone.0083700
- Murphy, R. O., & Ackerman, K. A. (2015). Social preferences, positive expectations, and trust based cooperation. *Journal of Mathematical Psychology*, *67*, 45–50. doi:10.1016/j.jmp.2015.06.001
- Myers-Levy, J., & Loken, B. (2015). Revisiting gender differences: What we know and what lies ahead. *Journal of Consumer Psychology*, *25*(1), 129–149. doi:10.1016/j.jcps.2014.06.003
- Niederle, M., & Vesterlund, L. (2007). Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics*, *122*(3), 1067–1101. doi:10.1162/qjec.122.3.1067

- Normann, H.-T., & Wallace, B. (2012). The impact of the termination rule on cooperation in a prisoner's dilemma experiment. *International Journal of Game Theory*, *41*(3), 707–718. doi:10.1007/s00182-012-0341-y
- Pansini, R., Shi, L., & Wang, R.-W. (2016). Women tend to defect in a social dilemma game in southwest China. *PLOS ONE*, *11*(11), e0166101. doi:10.1371/journal.pone.0166101
- Press, W. H., & Dyson, F. J. (2012). Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences of the United States of America*, *109*(26), 10409–10413. doi:10.1073/pnas.1206569109
- Pruitt, D. G., & Kimmel, M. J. (1977). Twenty years of experimental gaming: Critique, synthesis, and suggestions for the future. *Annual Review of Psychology*, *28*, 363–392. doi:10.1146/annurev.ps.28.020177.002051
- Pulford, B. D., Colman, A. M., & Loomes, G. (2018). Incentive magnitude effects in experimental games: Bigger is not necessarily better. *Games*, *9*(1), 4. doi:10.3390/g9010004
- Rapoport, A. (1967). A note on the “index of cooperation” for Prisoner's Dilemma. *Journal of Conflict Resolution*, *11*(1), 100–103. doi:10.1177/002200276701100108
- Rapoport, A., & Chammah, A. M. (1965a). *Prisoner's Dilemma: A study in conflict and cooperation*. Ann Arbor, MI: University of Michigan Press.
- Rapoport, A., & Chammah, A. M. (1965b). Sex differences in factors contributing to the level of cooperation in the Prisoner's Dilemma game. *Journal of Personality and Social Psychology*, *2*(6), 831–838. doi:10.1037/h0022678
- Rapoport, A. & Chammah, A. M. (1969). The game of Chicken. *American Behavioral Scientist*, *10*(3), 10–28. doi:10.1177/000276426601000303
- Rapoport, A., Seale, D. A., & Colman, A. M. (2015). Is tit-for-tat the answer? On the conclusions drawn from Axelrod's tournaments. *PLOS ONE*, *10*(7), 1–11, e0134128. doi:10.1371/journal.pone.0134128.
- Roth, A. E. (1995). Introduction to experimental economics. In J. Kagel & A. E. Roth (Eds.), *Handbook of experimental economics* (pp. 3–109). Princeton, NJ: Princeton University Press.
- Roth, A. E., & Murnighan, J. K. (1978). Equilibrium behavior and repeated play of the Prisoner's Dilemma. *Journal of Mathematical Psychology*, *17*(2), 189–198. doi:10.1016/0022-2496(78)90030-5
- Scodel, A., Minas, J. S., Ratoosh, P., & Lipetz, M. (1959). Some descriptive aspects of two-person non-zero-sum games. *Journal of Conflict Resolution*, *3*(2), 114–119. doi:10.1177/002200275900300203
- Selten, R., & Stoecker, R. (1986). End behavior in sequences of finite Prisoner's Dilemma supergames: A learning theory approach. *Journal of Economic Behavior and Organization*, *7*(1), 47–70. doi:10.1016/0167-2681(86)9002-1
- Simpson, B. (2003). Sex, fear, and greed: A social dilemma analysis of gender and cooperation. *Social Forces*, *82*(1), 35–52. doi:10.1353/sof.2003.0081
- Sobel, J. H. (1993). Backward-induction arguments: A paradox regained. *Philosophy of Science*, *60*(1), 114–133. doi:10.1086/289721
- Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review*, *81*, 971–978. doi:10.1007/bf01669298

- Tucker, A. (1950/2001). A two-person dilemma. Unpublished notes, Stanford University. Reprinted in E. Rasmussen (Ed.), *Readings in games and information* (pp. 7–8). Malden, MA: Blackwell.
- Wood, W., & Eagly, A. H. (2015). Two traditions in research on gender identity. *Sex Roles*, 73(11), 461–473. doi:10.1007/s11199-015-0480-2
- Yaffee, R., & McGee, M. (2000). *Introduction to time series analysis and forecasting with applications of SAS and SPSS*. Cambridge, MA: Academic Press.
- Yanovitzky, I., & VanLear, A. (2008). Time series analysis: Traditional and contemporary approaches. In A. F. Hayes, M. D. Slater, & L. B. Snyder (Eds.), *The SAGE sourcebook of advanced data analysis methods for communications research* (pp. 89–124). Thousand Oaks, CA: Sage Publications. doi:0.4135/9781452272054.n4