In Press, *Decision*, 2016, APA, DOI: 10.1037/dec0000057

Reasons for Cooperating in Repeated Interactions:

Social Value Orientations, Fuzzy Traces, Reciprocity, and Activity Bias

Briony D. Pulford, Andrew M. Colman, Catherine L. Lawrence, and Eva M. Krockow

University of Leicester

Author Note

Briony D. Pulford, Andrew M. Colman, Catherine L. Lawrence, and Eva M. Krockow, Department of Neuroscience, Psychology and Behaviour, University of Leicester.

Catherine L. Lawrence is now at the School of Psychology, Bangor University.

We are grateful to the Leicester Judgment and Decision Making Endowment Fund (Grant RM43G0176) for support in the preparation of this article, to Kenneth Malmberg and an anonymous referee for helpful comments, and to Kevin McCracken for developing the software used in the experiments.

Correspondence concerning this article should be addressed to Briony D. Pulford, Department of Neuroscience, Psychology and Behaviour, University of Leicester, Leicester LE1 7RH, United Kingdom. E-mail: bdp5@le.ac.uk

Abstract

Many human interactions involve patterns of turn-taking cooperation that can be modeled by the deeply paradoxical Centipede game. A backward induction argument suggests that cooperation is irrational in such interactions, but experiments have demonstrated that players cooperate frequently and earn better payoffs as a consequence. We formulate six competing theories of cooperation in Centipede games and report the results of 2 experiments, based on investigations of several closely matched games with different payoff structures and different methods of reaching decisions. The results show that turn-taking cooperation does not appear to be explained by reciprocity theory, activity bias theory, or a motive to maximize relative payoffs, but that collective rationality, in the form of a motive to maximize joint payoffs, and fuzzy-trace theory can explain cooperation in interactions of this type. Reciprocity increases cooperation across repeated games between fixed player pairs, but there is no evidence of reciprocity influencing cooperation within games.

Keywords: Centipede game, fuzzy-trace theory, reciprocity, social value orientation, team reasoning

Reasons for Cooperating in Repeated Interactions:

Social Value Orientations, Fuzzy Traces, Reciprocity, and Activity Bias

Human relationships afford many opportunities for turn-taking sequences of cooperative actions. This familiar feature of human social life is recognized in expressions such as "You scratch my back and I'll scratch yours" and "One good turn deserves another." Everyday examples include business associates taking turns writing references for each other's job applications, next-door neighbors taking turns looking after each other's pets, infantry soldiers taking turns providing covering fire for each other while advancing across disputed territory, house burglars taking turns acting as lookouts while their partners in crime steal property, and university professors taking turns reading and providing feedback on each other's research grant applications. Typically, the benefit b to the recipient of each cooperative action is greater, or at least no smaller, than the cost c to the cooperator, so that $c \leq b$. In this article, we formulate several competing theories that might explain this ubiquitous and important form of behavior, and we report the results of empirical tests of these theories, using Centipede games.

The strategic structure of reciprocal interactions of this type is depicted in Figure 1. This figure represents an eight-legged Centipede game (whose length could be extended indefinitely) involving two interactive decision makers or *players*, labeled A and B. The potential sequence of alternating decisions or *moves* is shown in the eight numbered *decision nodes* along the top row, and the players' *payoffs* (accumulated gains or losses that are realized only when the game ends) are shown in the *terminal nodes* in the bottom two rows and at the extreme right, Player A's payoffs stacked above Player B's. Player A begins at the first decision node on the left by choosing either STOP or GO. If Player A stops the game immediately, then the payoffs are 8 units to Player A and 6 units to Player B. If Player A chooses GO, then this hands the decision to Player B, who has a choice of either STOP or GO at the second decision node. At this second decision node, a STOP move yields accumulated payoffs of 6 units to Player A and 11 to B. The game continues in this way, and if neither player ever chooses STOP, then it terminates automatically after the eighth decision node, with accumulated payoffs of 20 units to Player A and 18 to B.

A STOP move merely ends the game with a relatively favorable payoff to the player who chooses STOP. A GO move keeps the game going and, in this particular case, always costs the player who makes it two units and benefits the co-player five units, so we have $c = 2$ and $b = 5$. The game models reciprocal turn-taking involving cooperation and defection, with GO moves generally representing cooperation, because they increase the joint payoff of the player pair at a cost to the GO-chooser, and STOP moves representing defection (noncooperation).

Ever since the Centipede game was introduced (Rosenthal, 1981), its paradoxical character has intrigued decision scientists. The paradox is revealed by the following logical argument, showing why strictly rational decision makers would never cooperate. The argument begins with three standard game-theoretic assumptions: (a) Both players are instrumentally rational in the sense that, whenever they face a choice between two options, knowing that one yields a higher payoff than the other, they choose the higher-paying option. If the payoffs represent their true preferences, then this means nothing more than that people choose the options they prefer. (b) Both players know the specification of the game—the information shown in Figure 1—and anything that can be logically deduced from it. (c) Assumptions (a) and (b) are *common knowledge*, in the sense that both players know them, know that both know them, know that both know that both know them, and so on for as many iterations as required (see below).

The *backward induction* argument proceeds as follows. If the eighth decision node were to be reached, then Player B would choose STOP, because that would earn a payoff of

20 units, whereas cooperating would earn only 18. Player B's rationality, defined in assumption (a) above, and knowledge of the payoffs (b) suffice to ensure this decision. If the seventh decision node were to be reached, then it follows that Player A would defect, because that would earn Player A 17 units, whereas cooperating would earn only 15, because (a) and (b) plus just the first iteration of (c) imply that Player A knows that Player B would defect at the next opportunity. This argument unfolds step by step, with an additional recursive clause of common knowledge (c) required at each successive step, until we reach the first decision node on the left, where Player A would defect, earning a payoff of 8 units, because (a), (b), and seven iterations of (c) logically imply that cooperation would yield a payoff of 6 units at the second decision node, where Player A knows that Player B would certainly defect. Common knowledge can be understood intuitively without the cognitive burden of working through all the individual steps. For example, a public announcement in a room full of people immediately becomes common knowledge among the people present (Milgrom, 1981).

The paradoxical character of the game is now clear. According to the backward induction argument, instrumentally rational players, who by definition invariably choose the right moves to maximize their own individual payoffs, earn only small payoffs, but irrational players earn much more by cooperating. In game-theoretic terminology, defecting at the first decision node is the optimal (subgame perfect) *Nash equilibrium* of the Centipede game (Busemeyer & Pleskac, 2009; Nash, 1950, 1951; Rapoport, 2003). The backward induction argument has been endorsed as mathematically sound by Aumann (1995, 1998), among others, but even people who follow the White Queen in *Through the Looking-Glass* in believing "as many as six impossible things before breakfast" (Carroll, 1871, chap. 5) might want to fortify themselves with a mug of strong coffee before swallowing it.

The Centipede game shown in Figure 1 has a joint payoff function that increases linearly, because the accumulated sum of the payoff pairs increases by 3 units following each cooperative move: 14, 17, 20, 23, …, up to 38 on the right. In some social interactions, payoffs accumulate at an accelerating rate—for example, revisiting the example of two business associates who take turns writing references for each other's job applications, perhaps each successive job is more important and better paid than the last, requiring a correspondingly longer and more detailed reference. Interactions with such accelerating payoffs are modeled by Centipede games with exponentially increasing payoff sums. In contrast to this there are classic bargaining interactions involving decisions that serve only to change the way a fixed payoff is divided, and such interactions can be modeled by constant-sum Centipede games. Figure 2 shows an exponential Centipede game in which the payoff sums are 2, 4, 8, 16, …, 512 and a constant-sum Centipede game in which the payoff sum remains fixed at 400. In constant-sum Centipede games, GO moves are not strictly cooperative, because they do not increase the joint payoff of the player pair (a standard definition of cooperation).

If the cost of a single cooperative move is c and the benefit to the recipient b, then in Figure 1, $c = 2$ and $b = 5$. If we further label a player's own payoff x and the co-player's payoff y, then, in Figure 2, the exponential Centipede game (a) has $c = x/3$ and $b = 6y$, and the constant-sum Centipede game (b) has $b = c = x - y + 25$ and $x + y = 400$. The exponential Centipede game is the canonical version most frequently discussed in the literature, but if the backward induction argument is valid, then all versions are at least as paradoxical as the better known Prisoner's Dilemma game. Instrumental rationality is defined in terms of maximizing individual payoffs, yet in Centipede games, irrationally cooperative players earn better payoffs than rational players. Cooperative players could legitimately ask their rational colleagues the familiar question: "If you're so clever, how come you ain't rich?"

Experimental investigations of Centipede games have reported rampant cooperation. The earliest experiments reported that only a tiny minority of players defected at the first decision node, and a substantial minority cooperated even at the last (Bornstein, Kugler, & Ziegelmeyer, 2004; Fey, McKelvey, & Palfrey, 1998; McKelvey & Palfrey 1992; Nagel & Tang, 1998, Parco, Rapoport, & Stein, 2002). Subsequent experiments have generally corroborated these findings, thereby providing vivid instances of the way in which normative theory and psychological facts can pass each other by (Johnson-Laird & Shafir, 1993).

The rest of this article is organized as follows: Next, we formulate and comment on several theories that appear to explain cooperation in Centipede games. Then we report an experiment designed to test theories based on social value orientations, fuzzy-trace theory, and reciprocity. Following that, we report a second experiment designed to test theories based on reciprocity (in a different way) and activity bias. In the final section, we draw the threads together and formulate some general conclusions.

Theories of Cooperation

Whatever the reason players cooperate in Centipede games, it cannot be to maximize their individual payoffs, knowing that co-players are doing likewise, because if they followed such game-theoretic reasoning, then they would never cooperate. Why, then, do they cooperate? The aim of this article is to formulate several new theories and to test them experimentally, but first it will be useful to clear the ground by commenting on a plausible suggestion frequently put forward by people when they first encounter the problem. This is the notion that players do not understand the game and therefore make arbitrary or random moves. Bearing in mind that game theory specifies only one rational way of playing the game—defecting at the first decision node—and many irrational ways, this implies that cooperation can be explained away as random fumbling in the dark.

This explanation overlooks the compelling intuitive appeal of cooperation. For example, the authors of this article understand the Centipede game, and the fact that we do is common knowledge among us, but if we played the game shown in Figure 1 with each other, we would all cooperate up to the seventh decision node at least. We are happy to note that we are in excellent company. The Nobel prizewinning game theorist Reinhard Selten, discussing a closely related game, said that he would not follow backward induction:

> From my discussions with friends and colleagues, I get the impression that most people share this inclination. In fact, up to now I met nobody who said that he would behave according to [backward] induction theory. My experience suggests that mathematically trained persons recognize the logical validity of the induction argument, but they refuse to accept it as a guide to practical behavior. (Selten, 1978, pp. 132–133)

Furthermore, if random fumbling were a major reason for cooperation, then we should expect a sharp decrease in cooperative choices when the game is repeated over many rounds as players' understanding surely increases through experience, but experimental evidence shows only a small decrease or none at all (e.g., El-Gamal, McKelvey, & Palfrey, 1993; Fey, McKelvey, & Palfrey, 1996; McKelvey & Palfrey, 1992; Parco, Rapoport, & Stein, 2002).

A more subtle and complex theory, based on bounded rationality rather than total incomprehension, is the *agent quantal response equilibrium* of McKelvey and Palfrey (1998). According to this theory, players make errors, and an equilibrium can be shown to result if they implement best replies to their co-players' moves imperfectly, with errors that are less likely the costlier they are. This and other model-fitting theories (Kawagoe & Takizawa, 2012; Rapoport, Stein, Parco, & Nicholas, 2003; Zauner, 1999) do an excellent job of describing and predicting behavior, but they do not even attempt to explain it psychologically.

The experiments described later in this article do not include tests of the random fumbling theory or formal model-fitting theories. We turn now to the most prominent cognitive and other psychological processes that appear to be capable of explaining cooperation in Centipede games and that are tested in the experiments that follow. These theories are prime examples of what Chater (2015) characterized as "cognitively informed rational models."

Social Value Orientation Theories

Possible reasons for cooperating include motives to maximize relative payoffs (to beat the co-player by the largest possible margin) or joint payoffs (to earn as much as possible for the player pair, considered as a unit) rather than individual payoffs, as assumed by default in orthodox game theory. Relative payoff maximization and joint payoff maximization are usefully conceptualized in terms of the psychological concept of *social value orientation* (SVO), introduced by Messick and McClintock (1968) and McClintock (1972) to model selfishness and other-regarding motivations, including relative and joint payoff maximization (for reviews, see Balliet, Parks, & Joireman, 2009; Bogaert, Boone, & Declerck, 2008; Rusbult & Van Lange, 2003; Van Lange, 2000). Early research focused on the *individualistic* SVO (maximizing individual payoffs), the *competitive* SVO (maximizing relative payoffs), and the *cooperative* SVO (maximizing joint payoffs). Subsequently, the *equality-seeking* SVO (minimizing the difference between payoffs) and the *altruistic* SVO (maximizing the co-player's payoff) were added to the theory, and attention focused on a hybrid *prosocial* SVO combining the cooperative and equality-seeking orientations (e.g., Van Lange 1999; Yamagishi et al., 2013). Approximately 57% of people are predominantly cooperative, 27% predominantly individualistic, and 16% predominantly competitive (Au & Kwong, 2004), and SVO correlates significantly with personality descriptions given by friends and roommates (Bem & Lord, 1979) and predicts everyday activities, including volunteering for charitable causes (McClintock & Allison, 1989; Van Lange, Bekkers, Schuyt, & Van Vugt, 2007). Messick and McClintock originally conceived of SVOs as an environmentally or circumstantially determined variable (what we call state SVO), and the earliest experiments explicitly ignored stable individual differences (trait SVO), although later research tended to focus on trait SVO and ignore state SVO.

In two-player games, SVO can be formalized as follows (Colman, Körner, Musy, & Tazdaït, 2011). Utilities representing players' true subjective preferences, revealed by their actual choices, are represented by the symbol U. Utilities that incorporate other-regarding preferences are interpreted as functions of the objective payoffs V of players and their co-players. The payoffs that are presented to participants in experimental games are objective payoffs, and as such they do not take other-regarding preferences into account. The key payoff transformations may now be defined as follows. Let v_i and v_j be the objective payoffs to Players i and j in a two-player game, and let s_i and s_j be their chosen strategies. Player i is assumed to maximize a utility function $U_i(s_i, s_j) = f_i(v_i, v_j)$, and Player i's SVO is a property of the particular function f_i that reflects i's motivation at the time. The *individualistic* SVO is defined for Player i by $f_i = v_i$ (players simply maximize their own objective payoffs), the *competitive* SVO by $f_i = v_i - v_j$ (players maximize the difference between their own and their co-players' objective payoffs), the *cooperative* SVO by $f_i = v_i + v_j$ (players maximize the sum of both players' objective payoffs), the *altruistic* SVO by $f_i = v_j$ (players maximize their co-players' objective payoffs), and the *equality-seeking* SVO by $f_i = -|v_i - v_j|$ (players minimize the difference between the objective payoffs). Players are invariably motivated to maximize their expected utilities U, as required by von Neumann and Morgenstern's (1947) utility theory (Busemeyer, 2015), but these expected utilities may be individualistic, competitive, cooperative, altruistic, or equality-seeking, depending on individual differences between players and circumstances of the social interaction.

Competitive SVO. One obvious possibility is that cooperative moves in Centipede games are motivated by the competitive SVO. This may appear self-contradictory, but players may make cooperative moves, not through any desire to cooperate per se, but on the contrary in a competitive attempt to maximize relative payoffs—hoping to beat their co-players by as large a margin as possible. In the exponential version in Figure 2 (top), Player A can defect at the first decision node and earn one unit more than Player B (earning 1.5 units compared to Player B's 0.5); but by cooperating, Player A might hope to earn four units more than Player B by defecting at the third decision node, or to beat Player B by a larger margin by defecting even later in the game. This makes sense in the exponential Centipede game, and indeed in the constant-sum game (Figure 2, bottom), but it has no persuasive force in the linear Centipede game shown in Figure 1, because in that version the differences between the players' objective payoffs do not increase on successive cooperative moves. This suggests an experimental test of this theory that could falsify it decisively if it is wrong, by determining whether players cooperate in games with constant or decreasing payoff differences, and this idea is incorporated in Experiment 1, described below.

Cooperative SVO. A closely related theory postulates a form of *collective rationality* captured by the cooperative SVO. Perhaps players aim to maximize *joint* payoffs—seeking to make the sum of payoffs to the player pair as large as possible. Collective rationality underlies theories of team reasoning (Bacharach, 1999, 2006; Sugden, 1993, 2005), and experimental evidence has confirmed that this occurs in suitable games (Bardsley, Mehta, Starmer, & Sugden, 2010; Butler, 2012; Colman, Pulford, & Lawrence, 2014; Colman, Pulford, & Rose, 2008). In all versions of the Centipede game apart from the constant-sum version, the sum of payoffs to both players increases as the game progresses, but in the constant-sum version, by definition, it remains fixed. This suggests an obvious experimental test of cooperative SVO theory, namely comparing cooperation in constant-sum Centipede games with cooperation in versions with other payoff functions. Experiment 1, described below, includes this test.

Equality-seeking SVO. In most Centipede games, inequality between the players' payoffs increases as successive cooperative moves are made. In both games in Figure 2, for example, the payoff difference is smaller after early than later defections. However, there are Centipede games in which payoff differences remain constant, as in Figure 1, and others in which payoff differences decrease. In view of experimental evidence that equality-seeking (or inequality aversion) has a powerful and pervasive effect on decision making in laboratory experiments and everyday life (Fehr & Schmidt, 1999; van den Bos et al., 2011; Van Lange 1999; Van Lange, De Bruin, Otten, & Joireman, 1997), Experiment 1, described below, includes an investigation of the equality-seeking motivation.

Fuzzy-Trace Theory

According to fuzzy-trace theory (Reyna & Brainerd, 1991, 1995, 2008, 2011), decision makers faced with problems such as the Centipede game generally encode and store two types of mental representations in parallel: *verbatim representations*, including precise numerical details, and *gist representations*, based on the most basic level of measurement that enables a decision to be made. A gist representation may be a *categorical gist* (some/none) or, if that fails to yield a determinate decision, then an *ordinal gist* (more/less, larger/smaller, some/more, and so on). Most adult decision makers prefer gist representations for reasoning and decision making, because they simplify and clarify reasoning tasks, and also because forgetting rates are higher for verbatim traces.

A possible explanation of cooperation in the Centipede falls out naturally from fuzzy-trace theory, in three parts. First, the most basic gist for players at the beginning of the versions in Figure 1 and Figure 2 (top) is the ordinal *some/more* gist: *STOP → certainty of small payoff* and *GO → possibility of larger payoff*. This emerges from even the most cursory glance at the game tree and, because players assumedly prefer larger to smaller payoffs, the theory predicts that they will therefore cooperate. Second, in linear Centipede games such as Figure 1, the same interpretation applies at later decision nodes, until the end of the game approaches, when *STOP → certainty of small payoff* no longer holds; hence fuzzy-trace theory predicts that players will tend to exit before the end of the game. Third, if we assume that players judge the magnitude of the possible payoff resulting from cooperation by estimating the average of the possible resulting payoffs, then this gist-based reason for cooperating seems less compelling in the constant-sum version shown in Figure 2 (bottom). Without calculation, Player A may not judge the average payoff following a GO move at the first decision node (actually 213) to be appreciably larger than the payoff resulting from a STOP move (200). For Player B, the average payoff following a GO move at the second decision node (182) is actually smaller than the payoff resulting from a STOP move (225). At later decision nodes, this is true for both players. Fuzzy-trace theory therefore suggests earlier defection in constant-sum versions, and this will also be tested in Experiment 1, described below.

Reciprocity Theory

Perhaps players are motivated to cooperate in Centipede games by considerations of reciprocity. The concept of *reciprocal altruism* was introduced by Trivers (1971) to explain the evolution of cooperation. He illustrated his idea with a hypothetical scenario in which a swimmer A with cramp has a 50-50 chance of drowning unless a passer-by B jumps in and attempts a rescue, in which case there is a 1 in 20 chance that both A and B will drown. If B believes that the situation will be repeated at some later date with the roles reversed, and that A will return the favor only if B helps on this occasion, then, by helping on this occasion, B trades a 50-50 chance of drowning at the later date (if A does not help) for a 1 in 10 chance (a 1 in 20 chance of drowning while trying to save A, plus 1 in 20 chance when A tries to save

B). It is therefore rational for B to help A, in spite of the danger. It turns out to be worthwhile whenever the benefit/cost ratio $b/c > 1/p$, where p is the probability of a later opportunity for cooperation arising with roles reversed (Nowak, 2006). When this probability is a certainty, as in Trivers's example, $p = 1$ and therefore $1/p = 1$, and it pays to help whenever the cost to the cooperator is less than the benefit to the recipient, that is, whenever $b/c > 1$, and hence whenever $c < b$.

Reciprocity is recognized as one of the most important explanations for cooperation (Imhof & Nowak, 2010; Trivers, 2005). Whether or not it provides an adequate explanation of cooperation in Centipede games can be tested by comparing how the game is played in its usual *extensive* form, in which players move sequentially, with how it is played in the so-called *normal form* (not the usual form for this game) in which both players simply specify simultaneously at which decision node they would defect. A normal-form Centipede game provides no opportunity for reciprocity, because players make single, simultaneous decisions, and there is no opportunity for returning favors. Any cooperation that is observed in such circumstances must have some other explanation. Experiment 2, reported below, includes the relevant comparison.

Activity Bias Theory

It is possible that players cooperate in Centipede games because of a bias in favor of doing something active rather than stopping the game. This may initially seem unlikely, because an opposite phenomenon, called *omission bias*, is well established in the judgment and decision making literature. Kahneman and Miller (1986) argued that commissions should lead to greater regret than omissions in decisions with possible negative outcomes, and Ritov and Baron (1990) showed that people are indeed reluctant to vaccinate children when the vaccination itself can cause death, even when they know that death from the vaccination is much less likely than death from the disease. However, Landman (1987) and Gleicher, Kost, Baker, Strathman, Richman, and Sherman (1990) found evidence for action bias in decisions involving positive outcomes.

We use the term *activity bias* to denote something slightly different: a more specific tendency in Centipede games to prefer activity (cooperation, allowing the game to continue) to inactivity (defection, stopping the game), and we hypothesize three reasons for it. First, cooperation is necessary to enable the co-player to participate and earn financial rewards, whereas defecting effectively shuts out the co-player and may therefore be considered impolite. Second, the *demand characteristics* of experiments, first adumbrated by Orne (1962), implicitly require activity from participants, and terminating an experiment prematurely may be interpreted as spoiling it, refusing to engage with the task, or behaving foolishly. Third, there is evidence that experimental participants generally prefer activity to inactivity (Wilson et al., 2014).

Activity bias theory, like reciprocity theory, lacks traction when the Centipede game is played in normal form, with each player making just a single strategy decision to cover all contingencies, because this provides no scope for further activity. A comparison of cooperation in normal-form and extensive-form Centipede games should reveal whether this theory has any validity, and relevant experimental evidence is reported in Experiment 2 below.

Next, we report an experiment designed to test SVO, fuzzy-trace, and reciprocity theories. This will be followed by a second experiment, carried out simultaneously with the same participant pool, involving a comparison with one of the treatment conditions in Experiment 1, designed to test reciprocity theory more directly and also to test activity bias theory.

Experiment 1

The aim of this experiment is to provide controlled tests of social value orientation theories, especially competitive SVO theory and cooperative SVO theory, and also fuzzy-trace theory and reciprocity theory. This is achieved by comparing cooperation in Centipede games with subtly different payoff structures. Theories of reciprocity and activity bias are tested in Experiment 2.

Method

Participants. The participants were 186 students and employees at the University of Leicester (112 female, 74 male), aged 18–53 years (*M* = 22.21, *SD* = 5.24) recruited via an electronic bulletin board at the university. The sample size was set in advance. Around 60% were British, 7% other European, 13% Chinese, and the rest were a diverse spread from around the world. Participants were remunerated according to a between-subjects version of the random lottery incentive system, a technique that successfully avoids problems associated with other payment schemes (Lee, 2008) and has been shown empirically to elicit true preferences (Cubitt, Starmer, & Sugden, 1998; Starmer & Sugden, 1991). We paid each participant a show-up fee of £5.00 (about $8.00) and entered every participant into a lottery in which three winners were paid an additional amount, up to a theoretical maximum of £84 ($134), according to their average payoffs across 20 repetitions of the Centipede game that they played in the experiment. In fact, the three lottery winners received £39.55, £40.75, and £52.30.

Materials. The basic materials included four linear Centipede games with different payoff structures (see Figure 3): a constant payoff-difference game, a constant-sum game, an increasing payoff-difference game, and a decreasing payoff-difference game. We did not use an exponential Centipede game, because it would then have been impossible to match it with other games required for our experimental purposes. For optimal experimental control, the four Centipede games that we devised are as similar as possible to one another, apart from certain differences in their payoff structures. In particular, they have identical numbers of decision nodes and similar payoff ranges, and they are all linear Centipede games with payoff sums satisfying $y = ax + b$, where y is the sum of Player A's and Player B's payoffs at a given decision node, x is the decision node number, and a and b are constant parameters defining particular linear payoff functions. For Game (a) in Figure 3, $a = 10$ and $b = 40$; for Game (b), $a = 0$ and $b = 104$; for Game (c), $a = 6$ and $b = 54$; and for Game (d), $a = 15$ and $b = 27$. For each game, two versions of the instructions were prepared to take account of different pairing conditions: anonymous fixed pairing of players throughout the 20 rounds, or anonymous random pairing after each round. Random pairing removes the possibility that a player may cooperate on one round with the conscious or unconscious aim of influencing the co-player to cooperate on the next.

To measure social value orientation, we asked all participants to fill in the Triple-Dominance Measure of Social Values (Van Lange, De Bruin, Otten, & Joireman, 1997). This is a short (nine-item) questionnaire-based measure of trait social value orientation.

Design. The experimental design was a between-subjects, two-factor, Game × Pairing factorial design. There were no additional unreported conditions. Participants were assigned randomly to one of the four games (constant payoff-difference, constant-sum, increasing payoff-difference, decreasing payoff-difference) and to one of the two pairing conditions, anonymous fixed pairing or anonymous random pairing (perfect stranger matching), and they remained in the same treatment conditions and in the same role (Player A or Player B) throughout the 20 rounds of the game. Under both fixed and random pairing, participants were unaware of the identity of their co-players. The principal dependent variable was the mean exit node across all player pairs.

Procedure. The experiment was conducted during 12 one-hour testing sessions, with approximately 14 to 24 participants per session in a single very large laboratory, and was controlled with custom software. The participants were randomly assigned to different treatment conditions within each testing session. After signing a consent form, each participant sat in front of a networked computer monitor that displayed the appropriate game for the relevant treatment condition. The game was displayed diagrammatically, as in Figure 1, but with decision nodes and payoffs in different colors for Player A and Player B. Textual instructions, accompanied by examples, explained the rules of the game, the payoff functions, the number of rounds to be played, and the incentive scheme. Each participant was also given a leaflet containing a diagram of the relevant Centipede game and a summary of the instructions.

The wording of the instructions was identical across treatment conditions, apart from information relating to specific payoffs that varied between treatment conditions and instructions regarding pairing after successive rounds: "You will be paired with the same other participant each time" or "You will be randomly paired with a different participant each time." The on-screen instructions concluded as follows:

> At the end of the experiment you will be entered into a lottery. Three people will have their average payoff across all decisions converted to cash. Therefore, it is in your interest to consider the payoffs and choose between GO or STOP carefully, as this will determine how much cash you could be paid.

The experimenter showed the participants a large pile of money to help convince them of the reality of the incentives, and invited them to seek clarification about anything they did not understand. The participants worked through the experiment, interacting via their terminals at their own pace, apart from a constraint imposed by the anonymous pairing treatment conditions that sometimes required players to wait for a new co-player to complete a previous round before beginning the next round. At the start of each round, participants were reminded on-screen of their role as Participant A or Participant B as well as the number (out of 20) of the current round, and the Centipede game diagram was accompanied by the following text: "Participant A can choose to click on GO or STOP at (1) on the diagram above. Participant A please make your decision now. Participant B please wait for the outcome." Whenever a player defected by clicking STOP, the on-screen text reminded both players of the payoffs to each. After completing 20 rounds, participants were asked to fill in the questionnaire designed to measure trait SVO, then they were thanked, and demographic information and email addresses were recorded. Once all testing sessions were complete, three participants were randomly selected and remunerated with the average of their payoffs across all 20 games.

Results[1]

Mean exit nodes. Figure 4 displays the proportion of games ending at each of the terminal nodes, in each of the four experimental games, under fixed or random player pairing, and Figure 5 shows the mean exit nodes per round across the four games under fixed and random pairing.

A two-way ANOVA, with player pairs as the units of statistical analysis, examined the effects of game and pairing factors on mean exit nodes. The mean exit node differed significantly across games: $F(3, 85) = 12.93$, $p < .001$, partial $\eta^2 = .31$. In the constant-sum condition the mean exit node was 2.82; in the constant payoff-difference condition it was 4.02; in the increasing payoff-difference condition it was 4.48; and in the decreasing payoff condition it was 5.80. Each of these means differs significantly from each of the others ($p < .02$, LSD test) apart from the difference between the increasing payoff-difference and the constant payoff-difference conditions. There was also significantly greater cooperation, resulting in a slightly later mean exit node, under fixed pairing ($M = 4.62$) than under random pairing (M = 3.94), and this difference is small but significant: $F(1, 85) = 3.94$, $p = .05$, partial $\eta^2 = .04$. There was no significant interaction between pairing condition and game ($F < 1$).

Time series.[2] Figure 6 shows the sequence plots of mean exit nodes over the 20 rounds of the experiment, under fixed pairing and random pairing, in each of the four experimental games. It is obvious by inspection that there was more cooperation under fixed than random pairing across almost all 20 rounds in all games except the decreasing payoff-difference game. We performed time series analysis (e.g., Bisgaard & Kulahci, 2011; Chatfield, 2003; Shumway & Stoffer, 2006; Yanovitzky & VanLear, 2008) to the mean exit nodes, recorded for each round, in each of our treatment conditions—four distinct games crossed with fixed or random pairing. Time series analysis is a dynamic method of analysis that provides estimates of the nature and strength of effects of earlier decisions on later decisions and of the form of the overall trend over rounds. We relied on the SPSS Expert Modeler to select the best-fitting models to the time series.

In the constant payoff-difference game under fixed pairing, the best-fitting model is an ARIMA(0, 0, 0) model, indicating a lack of temporal structure in the data; but under random pairing the best fit is provided by a Holt model, an exponential smoothing model that fits time series with autocorrelation and linear trend, confirming the significance of the decline in cooperation over rounds under fixed pairing that is apparent in the sequence plot in Figure 6. In the constant-sum game, once again, the best-fitting model under fixed pairing is an ARIMA(0, 0, 0) model, indicating a lack of temporal structure in the data, and under random pairing the best fit is a Holt model, confirming linear and statistically significant decline in cooperation over rounds. In the increasing payoff-difference game under both fixed and random pairing, a Holt model provides the best fit, indicating a significant linear decline in cooperation over rounds. In the decreasing payoff-difference game, the best fit under fixed pairing is a Holt model and under random pairing is an exponential smoothing Brown model—a simple one-parameter version of the Holt model, also indicative of linear trend. These results suggest that, under both fixed and random pairing, cooperation declined significantly and linearly over rounds in the decreasing payoff-difference game.

SVO and defection. We summed the number of times that each participant defected (out of 20 rounds) and correlated these sums with their SVO subscale scores (cooperative, individualistic, and competitive). To assess the significance of the correlations while controlling for multiple tests, we tested the correlations against a false discovery rate of $q^* = .05$ (Benjamini & Hochberg, 1995). The results are displayed in Table 1. Under random pairing, there is a significant correlation ($r = .61$) between competitive SVO subscale score and frequency of defection in the decreasing payoff-difference game, but not in any other game. All other correlations are non-significant. Analyses also revealed no significant influence of gender on how often participants defected; males defected 9.07 times on average and females 8.88, $t(184) = 0.30$, $p = .76$.

Discussion

According to competitive SVO theory, cooperative moves in Centipede games are motivated by players seeking to maximize the difference between their own payoffs and those of their co-players, by enticing their co-players to cooperate and then defecting against them later. If this were the sole reason for cooperation, then we should expect no cooperation at all in games with decreasing payoff differences, because in any such game, a player maximizes the payoff difference by defecting as early as possible. Our results revealed, on the contrary, that players defected significantly *later*, on average, in the game with decreasing payoff differences than in any other game (see Figures 4, 5, and 6). However, in the random pairing condition, players with high scores on the competitive SVO subscale tended to defect significantly more frequently than others in the game with decreasing payoff differences (see

Table 1). Hence, the high overall level of cooperation in the game with decreasing payoff differences must be due to the behavior of players with other types of SVO. Competitive SVO is clearly refuted as a general explanation of cooperation in Centipede games.

According to cooperative SVO theory, Centipede games with increasing payoff sums should elicit more cooperation than the constant-sum game that we included in our experiment. The evidence clearly confirms this prediction: significantly less cooperation occurred in the constant-sum game than in the other games used in the experiment, all of which have increasing payoff sums (see Figures 4, 5, and 6), and the differences are quite striking. The same effect is predicted by fuzzy-trace theory, although for an entirely different reason (see below).

In our other games, the constant-sum game and increasing payoff-difference game, cooperation increases payoff inequality, but each cooperative move provides a larger benefit to the co-player than the last. This may be attractive to cooperative players in particular and may motivate them to continue cooperating in those games in spite of the increasing payoff inequality that results, because cooperative players tend also to be altruistic.

The equality-seeking SVO on its own cannot explain cooperation in Centipede games in general, because in most such games cooperation does not cause payoffs to become more equal. The exception is the (unusual) decreasing payoff-difference game that we included in our experiment. Figure 5 shows that cooperation was indeed much higher in this game than in any other, and our data therefore provide further evidence for the widespread inequality aversion that has been noted by previous researchers (Fehr & Schmidt, 1999; van den Bos et al., 2011; Van Lange 1999; Van Lange, De Bruin, Otten, & Joireman, 1997). There is no other obvious explanation for this finding.

Fuzzy-trace theory predicts cooperation at the early decision nodes in all four versions of the Centipede game shown in Figure 3, because the most primitive gist that yields a determinate decision, the *some/more* ordinal gist, suggests that a STOP move leads to a small payoff and a GO move the possibility of a larger payoff, and it is clear from Figures 4, 5, and 6 that this prediction is confirmed. This theory also predicts that, in all four versions of the game (all linear Centipede games), players will tend to cooperate less at later decision nodes, because at later nodes the certain payoff from defection is no longer appreciably smaller than the expected payoff from cooperating. Figure 4 shows that this prediction is confirmed in the constant payoff-difference and constant-sum game, and generally confirmed from the third decision node onwards in the increasing payoff-difference game. The decreasing payoff-difference game appears to be a special case, with an unusual pattern of cooperation over decision nodes.

Fuzzy-trace theory also predicts less cooperation and earlier defection in the constant-sum version in Figure 3 than in the other versions, because Player A's average payoff from cooperation at the first decision node (53.5) may not be appreciably larger, without calculation, than the payoff that is certain to result from defection (52), and Player B's average payoff from cooperation at the second decision node (50) is less than the payoff from defection (55), with similar considerations at later nodes, whereas in the other games the average payoff from early cooperation is much larger than from early defection. The *some/more* ordinal gist thus leads to the prediction of less cooperation in the constant-sum version. It is evident from Figures 4, 5 and 6 that this prediction was also confirmed.

According to reciprocity theory, players cooperate in Centipede games in the hope that their co-players will respond reciprocally and that both will benefit from mutual cooperation in the long run. If this process plays a major part in explaining cooperation in general, then we should expect greater cooperation in the fixed pairing than the random pairing treatment conditions, because fixed pairing allows players to engage in reputation management, cooperating on one round in the hope of influencing their co-players' cooperativeness on a later round. That is exactly what we observed, except in the decreasing payoff-difference game (see Figure 5). It is worth commenting that this effect may have occurred without the conscious awareness of the players. The probable reason why no significant reciprocity appears to have been elicited in the decreasing payoff-difference game is that it was unnecessary. Equality-seeking is known from previous research to be a powerful motive in experimental games (e.g., Au & Kwong, 2004; Fehr & Schmidt, 1999; van den Bos et al., 2011; Van Lange 1999; Van Lange, De Bruin, Otten, & Joireman, 1997; Yamagishi et al., 2013), and this is the only one of our games in which cooperation has the incidental side-effect of increasing equality. Cooperation on early rounds was far higher than in any of the other games, and although it declined slightly over later rounds, it remained higher than in any other game (see Figures 4 and 6). We suggest that, in this decreasing payoff-difference game, even in fixed pairing conditions, players felt less need to encourage cooperation from their co-players through reciprocity, because they expected increasing equality to have this effect in any case.

Our data show clear evidence of between-game reciprocity in the other games (apart from the decreasing payoff-difference game), because there is no other obvious explanation for the significantly greater cooperation in the fixed than random pairing conditions. But this does not constitute evidence for reciprocity as a driver of cooperation *within* Centipede games, or in repeated plays under random pairing. Experiment 2 provides a stringent test of reciprocity theory as an explanation of cooperation within individual Centipede games.

Experiment 2

The main purpose of this experiment is to test reciprocity and activity bias theories of cooperation in Centipede games. This is achieved by comparing cooperation in the linear constant payoff-difference game (Figure 3, top), played in its usual extensive form, with cooperation in a normal-form version of the same game. Other factors that might influence cooperation were controlled as strictly as possible.

Method

Participants. The participants were 48 students and employees at the University of Leicester (30 female, 18 male), aged 18–53 years (*M* = 23.23, *SD* = 6.20), 24 of whom participated in Experiment 1 and were assigned to the constant payoff-difference treatment condition with random pairing. In this second experiment, conducted in the same testing sessions as Experiment 1 to maximize comparability of procedures and participant pools and thus to optimize experimental control, 24 additional players were given the task of responding to a normal-form version of the game. The incentive scheme was exactly as in Experiment 1.

Materials. The Centipede game used in this experiment was the constant payoff-difference game (Figure 3, top) and a normal-form version of the same game (strictly speaking, the *reduced normal form* version, because we did not require players to specify how they would choose in situations that could not arise in the game; for example, if they chose to defect at the first decision node, then they did not also have to specify how they would choose at the third). Nagel and Tang (1998) reported the results of an experiment using a task that was isomorphic with the (reduced) normal form of a Centipede game, although it did not actually resemble a Centipede game. Their study made no attempt to compare choices in the normal form to choices in an equivalent extensive form of the game. As far as we are aware, no previous experimental study has reported differences in behavior between normal and extensive forms.

Design. The experimental design was a randomized groups design. Participants were randomly assigned to treatment conditions involving the normal and extensive forms of the game. As in Experiment 1, the dependent variable was the mean exit node in each round.

Procedure. The procedure was as in the random pairing, constant payoff-difference treatment condition in Experiment 1, except that half the participants were randomly assigned to the normal form of the game. Instructions were as similar as possible in the two treatment conditions. In the normal-form condition, after reading the standard on-screen instructions explaining the Centipede game, with the same text and diagrams as in the extensive-form condition, participants were given the following special instructions:

> Instead of starting at the circle on the left and continuing until you or the other player chooses STOP, you and the other player will make a single decision on each decision sequence. This will save a lot of time in the testing session. You will consider the payoffs and choose the circle where you would STOP the decision sequence if the other player has not stopped it earlier. If you would not stop the decision sequence and would GO until the end then please click on the final GO in the sequence. . . . The computer will then take into account the circle number that you chose and the circle number that the other participant chose. Both you and the other participant will receive the payoffs at the circle with the lower number. You will be told the payoffs for this circle. You will then be randomly paired with a different participant and asked to make a decision at which circle to stop the decision sequence again. The decision sequence will be presented 20 times in total and you will be randomly paired with a different participant each time. You will never know who you are paired with.

After 20 rounds had been completed, demographic information and email addresses were recorded. In all other respects, the procedure was the same as in Experiment 1.

Results[1]

Mean exit nodes. Figure 7 shows the proportion of games ending at each terminal node in extensive form and normal form of the constant payoff-difference game with random player pairing. An analysis of mean exit nodes showed that the normal form elicited a similar mean exit node (M = 3.47) compared to the extensive form (M = 3.40), $t(22) = 0.16$, $p = .87$.

Time series.[2] We performed time series analysis once again on the mean exit nodes per round. Sequence plots of mean exit nodes in the normal-form and extensive-form treatment conditions are shown in Figure 8, from which it is immediately obvious that cooperation was similar in both conditions on almost every round of the game, from the first to the last. In both the normal form and the extensive form, a Holt model provides the best fit, confirming a linear decline in cooperation over rounds in both the extensive and normal forms of this game. It is worth commenting that Nagel and Tang (1998), who used a number-choosing task, found no decrease in cooperation over 100 rounds. This suggests that their task, although formally equivalent to a normal-form Centipede game, may not be psychologically equivalent to it. Our normal-form Centipede game resembles the extensive-form version as closely as possible.

Discussion

We found similar levels of cooperation in the game played in normal form and extensive form, with similar time series showing similar patterns of autocorrelation and linear decline over rounds. Both reciprocity and activity bias theories rely on repetitions to explain cooperation, and both therefore predict no cooperation in the normal form, or at least significantly less cooperation than the extensive form. Because the games were identical in all other respects and were presented to the players in the same way, the only difference being in how they played the game, this finding fails to corroborate both theories. This suggests that neither reciprocity nor activity bias can be a major driver of cooperation in the

Centipede game. Not only did a substantial level of cooperation occur in the normal-form treatment condition, but a similar level and pattern of cooperation was observed in this condition and in the extensive-form condition. In the normal-form condition, neither theory has any explanatory power, because the players have no scope for activity or reciprocal turn-taking. Our findings therefore provide no support for these theories, because both rely on the opportunities for repeated cooperative choices in the course of dynamic play to explain why players cooperate in Centipede games.

General Discussion and Conclusions

The findings of our experiments permit some unexpectedly decisive answers to be given to the question of why people cooperate in repeated sequences of reciprocal actions even when, according to standard game-theoretic assumptions, a backward induction argument suggests that it is irrational to do so. The answer appears to be that cooperation may be motivated, in large part at least, by a desire to maximize the joint payoff of both interacting participants considered as a unit, or cooperation may result from players forming *some/more* fuzzy-trace representations of the structure of Centipede games. Our findings provide no evidence to support any of the other theories. We found evidence for inequality aversion and reciprocity, but neither of these could explain the cooperative behavior observed in the games.

In terms of the conception of rationality that is generally accepted in decision theory and game theory, human decision makers are limited by bounded rationality (Simon, 1956, 1982), and, if the backward induction argument is valid, then cooperation in repeated interactions must be rooted in some form of irrationality, because according to that argument rational players never cooperate. Formal theories based on special assumptions regarding the precise nature of the bounded rationality make fairly accurate predictions of behavior in Centipede games (McKelvey & Palfrey 1998; Rapoport, Stein, Parco, & Nicholas, 2003;

Zauner, 1999). But it is one thing to model irrational decisions mathematically and quite another to specify why these decisions are made and what motivates them. In this article, we examined six psychological theories of cooperation in Centipede games, all of which offer plausible reasons why players might cooperate, and we found decisive experimental evidence against three of them.

We expected our experimental evidence to support reciprocity theory, partly because reciprocity is a well-established phenomenon in human interaction and partly because introspection seemed to suggest that the compelling urge to cooperate in Centipede games may be based on considerations of reciprocity. Reciprocity operated *between* (rather than within) some of the games in Experiment 1, in which each of our four experimental games was played by half the players under fixed pairing for 20 rounds and by the other half under random pairing after every round. In three of the four games, we observed significantly more cooperation under fixed than random pairing, persisting throughout virtually all 20 rounds, and the obvious explanation for this is reciprocity. Players in fixed pairing conditions must have cooperated partly in the hope or expectation that their goodwill would be reciprocated in later rounds. Nevertheless, reciprocity fails as an explanation of cooperation *within* individual games, because in Experiment 2, a similar level and pattern of cooperation occurred in the normal-form Centipede game, in which considerations of reciprocity do not arise, and in an equivalent extensive-form version of the same game, in which there was scope for reciprocity to operate. This finding also casts doubt on activity bias theory, because there is no opportunity for activity in the normal form, yet players cooperated just as much.

We tested two theories based on social value orientations in Experiment 1, and one of them fared badly. Our results show that the competitive SVO, according to which players cooperate to maximize the difference between their own and their co-players' payoffs by enticing them to cooperate and then exiting at later decision nodes, cannot explain cooperation in Centipede games, because we observed significantly higher levels of cooperation in the decreasing payoff-difference game than in constant payoff-difference and increasing payoff-difference games, and that is exactly the opposite of what is predicted by the theory. The cooperative SVO, according to which players cooperate to maximize the joint payoff of the player pair, was corroborated by our data. In Experiment 1, we observed significantly less cooperation in the constant-sum game than in any of the other three games. Bearing in mind that the experimental games were closely matched, apart from necessary differences in their payoff functions, the only obvious explanation for this striking effect seems to be the fact that the constant-sum game is the only one in which the joint payoff of the player pair is not increased by cooperation. In other respects, the games are quite similar. We conclude from this that cooperation in Centipede games may be explained, at least in part, by collective rationality in the form of joint payoff maximization. There is evidence for joint payoff maximization in games very different from the Centipede game (Bardsley, Mehta, Starmer, & Sugden, 2010; Butler, 2012; Colman, Pulford, & Rose, 2008; Colman, Pulford, & Lawrence, 2014). The experiments reported in this article add another, quite different class of interactions that appear to elicit collective rationality.

An alternative explanation for reciprocal cooperation in Centipede games that is also consistent with our data is supplied by fuzzy-trace theory. The *some/more* ordinal gist, the simplest that provides a player with a decisive way of choosing at the beginning of the game, is *STOP → certainty of small payoff* and *GO → possibility of larger payoff*. From this we can derive the predictions that players will cooperate at early decision nodes, that cooperation will decline at later decision nodes, and that cooperation will occur less frequently in constant-sum than increasing-sum Centipede games. All of these predictions were corroborated in Experiment 1, suggesting that gist representation can indeed explain our data. It is possible that cooperative SVO and fuzzy-trace theory may provide different parts of the explanation for cooperation in Centipede games; further research is required on this issue. It goes almost without saying that there may be other factors at work that we failed to include in our investigation. If so, then cooperation in Centipede games may be explained by more than joint payoff maximization and fuzzy-trace theory, but it seems impossible to avoid the conclusion that these two processes provide at least large parts of the explanation.

We found gentle declines in cooperation over rounds in most treatment conditions, some so slight that it was impossible to judge merely by inspection whether they were real, and time series analysis was necessary to establish their statistical significance. This pattern of results is in line with findings of previous research into two-player Centipede games, which have typically reported small declines or none at all. We also found evidence for inequality aversion, or the equality-seeking SVO, in Experiment 1. The decreasing payoff-difference game, the only one of our four experimental games in which cooperation causes payoff differences to decrease, elicited significantly more cooperation than any other, and it is difficult to think of any explanation for this finding other than inequality aversion. This phenomenon is well established in other areas of experimental research (Fehr & Schmidt, 1999; van den Bos et al., 2011; Van Lange, 1999; Van Lange, De Bruin, Otten, & Joireman, 1997) and, once again, our findings show the same mechanism operating in a different domain of social interaction.

The experiments reported in this article go some way toward explaining cooperation in repeated interactions with the strategic structure of the Centipede game. Cooperation appears to be explained by a motive to maximize the joint payoff of the player pair, or by a tendency for players to respond to a gist rather than a verbatim representation of the problem, or possibly by both of these processes operating in tandem. Direct reciprocity appears to operate between repetitions of the game with fixed player pairing, but reciprocity does not explain cooperation within conventional Centipede games played sequentially.

References

Au, W. T., & Kwong, J. Y. Y. (2004). Measurements and effects of social-value orientation in social dilemmas: A review. In R. Suleiman, D. V. Budescu, I. Fischer, & D. M. Messick (Eds.), *Contemporary Psychological Research on Social Dilemmas* (pp. 71–98). Cambridge, UK: Cambridge University Press.

Aumann, R. J. (1995). Backward induction and common knowledge of rationality. *Games and Economic Behavior, 8*, 6–19. doi: 10.1016/S0899-8256(05)80015-6

Aumann, R. J. (1998). On the Centipede game. *Games and Economic Behavior, 23,* 97–105. doi: 10.1006/game.1997.0605

Bacharach, M. (1999). Interactive team reasoning: A contribution to the theory of co-operation. *Research in Economics, 53,* 117–147. doi: 10.1006/reec.1999.0188

Bacharach, M. (2006). *Beyond individual choice: Teams and frames in game theory* (N. Gold, & R. Sugden, Eds.). Princeton, NJ: Princeton University Press.

Balliet, D., Parks, C., & Joireman, J. (2009). Social value orientation and cooperation in social dilemmas: A meta-analysis. *Group Processes & Intergroup Relations, 12,* 533–547. doi: 10.1177/1368430209105040

Bardsley, N., Mehta, J., Starmer, C., & Sugden, R. (2010). Explaining focal points: Cognitive hierarchy theory versus team reasoning. *Economic Journal, 120,* 40–79. doi: 10.1111/j.1468-0297.2009.02304.x

Bem, D. J., & Lord, C. G. (1979). Template matching: A proposal for probing the ecological validity of experimental settings in social psychology. *Journal of Personality and Social Psychology, 37,* 833–846. doi: 10.1037/0022-3514.37.6.833

Benjamini, Y. & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B, 85,* 289–300.

Bisgaard, S., & Kulahci, M. (2011). *Time series analysis and forecasting by example.* Hoboken, NJ: Wiley.

Bogaert, S., Boone, C., & Declerck, C. (2008). Social value orientation and cooperation in social dilemmas: A review and conceptual model. *British Journal of Social Psychology, 47,* 453–480. doi:10.1348/014466607X244970

Bornstein, G., Kugler, T., & Ziegelmeyer, A. (2004). Individual and group decisions in the centipede game: Are groups more “rational” players? *Journal of Experimental Social Psychology, 40,* 599–605. doi: 10.1016/j.jesp.2003.11.003

Busemeyer, J. R. (2015). Cognitive science contributions to decision science. *Cognition, 135,* 43–46. doi:10.1016/j.cognition.2014.11.010

Busemeyer, J. R., & Pleskac, T. J. (2009). Theoretical tools for understanding and aiding dynamic decision making. *Journal of Mathematical Psychology, 53,* 126–138. doi:10.1016/j.jmp.2008.12.007

Butler, D. J. (2012). A choice for “me” or for “us”? Using we-reasoning to predict cooperation and coordination in games. *Theory and Decision, 73,* 53–76. doi: 10.1007/s11238-011-9270-7

Carroll, L. (1871). *Through the looking-glass.* London, UK: Macmillan.

Chater, N. (2015). Can cognitive science create a cognitive economics? *Cognition, 135,* 52–55. doi: 10.1016/j.cognition.2014.10.015

Chatfield, C. (2003). *The analysis of time series: An introduction* (6th ed.). New York: Chapman & Hall/CRC.

Cohen, J. (1983). The cost of dichotomization. *Applied Psychological Measurement, 7,* 249–253. doi: 10.1177/014662168300700301

Colman, A. M., Körner, T. W., Musy, O., & Tazdaït, T. (2011). Mutual support in games: Some properties of Berge equilibria. *Journal of Mathematical Psychology, 55,* 166–175. doi: 10.1016/j.jmp.2011.02.001

Colman, A. M., Pulford, B. D., & Lawrence, C. L. (2014). Explaining strategic coordination: Cognitive hierarchy theory, strong Stackelberg reasoning, and team reasoning. *Decision, 1,* 35–58. doi: 10.1037/dec0000001

Colman, A. M., Pulford, B. D., & Rose, J. (2008). Collective rationality in interactive decisions: Evidence for team reasoning. *Acta Psychologica, 128,* 387–397. doi: 10.1016/j.actpsy.2007.08.003

Cubitt, R. P., Starmer, C., & Sugden, R. (1998). On the validity of the random lottery incentive system. *Experimental Economics, 1,* 115–131. doi: 10.1007/BF01669298

El-Gamal, M. A., McKelvey, R. D., & Palfrey, T. R. (1993). A Bayesian sequential experimental study of learning in games. *Journal of the American Statistical Association, 88,* 428–435. doi: 10.2307/2290321

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics, 114,* 817–868. doi: 10.1162/003355399556151

Fey, M., McKelvey, R. D., & Palfrey, T. R. (1996). An experimental study of constant-sum centipede games. *International Journal of Game Theory, 25,* 269–287. doi: 10.1007/BF02425258

Gleicher, F., Kost, K. A., Baker, S. M., Strathman, A. J., Richman, S. A., & Sherman, S. J. (1990). The role of counterfactual thinking in judgments of affect. *Personality and Social Psychology Bulletin, 16,* 284–295. doi: 10.1177/0146167290162009

Imhof, L. A., & Nowak, M. A. (2010). Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B, 277,* 463–468. doi: 10.1098/rspb.2009.1171

Johnson-Laird, P. N., & Shafir, E. (1993). The interaction between reasoning and decision making: An introduction. *Cognition, 49,* 1–9. doi: doi:10.1016/0010-0277(93)90033-R

Kahneman, D., & Miller, D. T. (1986). Norm theory: Comparing reality to its alternatives. *Psychological Review, 93,* 136–153. doi: 10.1037/0033-295X.93.2.136

Kawagoe, T., & Takizawa, H. (2012). Level-*k* analysis of experimental Centipede games. Journal of Economic Behavior & Organization, 82, 548–566. doi: 10.1016/j.jebo.2012.03.010

Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin, 13,* 524–536. doi: 10.1177/0146167287134009

Lee, J. (2008). The effect of the background risk in a simple chance improving decision model. *Journal of Risk and Uncertainty, 36,* 19–41. doi: 10.1007/s11166-007-9028-3

McClintock, C. G. (1972). Social motivation: A set of propositions. *Behavioral Science, 17,* 438–454. doi: 10.1002/bs.3830170505

McClintock, C. G., & Allison, S. T. (1989). Social value orientation and helping behavior. *Journal of Applied Social Psychology, 19,* 353–362. doi: 10.1111/j.1559-1816.1989.tb00060.x

McKelvey, R. D., & Palfrey, T. R. (1992). An experimental study of the Centipede game. *Econometrica, 60,* 803–836. doi: 10.2307/2951567

McKelvey, R., D., & Palfrey, T. R. (1998). Quantal response equilibria for extensive form games. *Experimental Economics, 1,* 9–41. doi:10.1007/BF01426213

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology, 4,* 1–25. doi: 10.1016/0022-1031(68)90046-2

Milgrom, P. (1981). An axiomatic characterization of common knowledge. *Econometrica, 49,* 219–222. doi: 10.2307/1911137

Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review, 18,* 13–41. doi: 10.1177/1088868313501745

Nagel, R., & Tang, F. F. (1998). Experimental results on the Centipede game in normal form: An investigation on learning. *Journal of Mathematical Psychology, 42,* 356–384. doi: 10.1006/jmps.1998.1225

Nash, J. F. (1950). Equilibrium points in n-person games. *Proceedings of the National Academy of Sciences of the USA, 36,* 48–49. doi: 10.2307/88031

Nash, J. F. (1951). Non-cooperative games. *Annals of Mathematics, 54,* 286–295. doi: 10.2307/1969529

Nowak, M. A. (2006). Five rules for the evolution of cooperation. *Science, 314,* 1560–1563. doi: 10.1126/science.1133755

Orne, M. T. (1962). On the social psychology of the psychological experiment: With particular reference to demand characteristics and their implications. *American Psychologist, 17*, 776–783. doi: 10.1037/h0043424

Parco, J. E., Rapoport, A., & Stein, W. E. (2002). Effects of financial incentives on the breakdown of mutual trust. *Psychological Science, 13,* 292–297. doi: 10.1111/1467-9280.00454

Rapoport, A. (2003). Games: Centipede. In L. Nadel (Ed.), Encyclopedia of cognitive science (Vol. 2, pp. 196–203). London, UK: Nature Publishing Group.

Rapoport, A., Stein, W. E., Parco, J. E., & Nicholas, T. E. (2003). Equilibrium play and adaptive learning in a three-person Centipede game. *Games and Economic Behavior, 43,* 239–265. doi: 10.1016/S0899-8256(03)00009-5

Reyna, V. F., & Brainerd, C. J. (1991). Fuzzy-trace theory and framing effects in choice: Gist extraction, truncation, and conversion. *Journal of Behavioral Decision Making, 4,* 249–262. doi: 10.1002/bdm.3960040403

Reyna, V. F., & Brainerd, C. J. (1995). Fuzzy-trace theory: An interim synthesis. *Learning and Individual Differences, 7,* 1–75. doi: 10.1016/1041-6080(95)90031-4

Reyna, V. F., & Brainerd, C. J. (2008). Numeracy, ratio bias, and denominator neglect in judgments of risk and probability. *Learning and Individual Differences, 18*, 89–107. doi:10.1016/j.lindif.2007.03.011

Reyna, V. F., & Brainerd, C. J. (2011). Dual processes in decision making and developmental neuroscience: A fuzzy-trace model. *Developmental Review, 31,* 180–206. doi: 10.1016/j.dr.2011.07.004

Ritov, I., & Baron, J. (1990). Reluctance to vaccinate: Omission bias and ambiguity. *Journal of Behavioral Decision Making, 3*, 263–277. doi: 10.1002/bdm.3960030404

Rosenthal, R. W. (1981). Games of perfect information, predatory pricing and chain-store paradox. *Journal of Economic Theory, 25*, 92–100. doi: 10.1016/0022-0531(81)90018-1

Rusbult, C. E., & Van Lange, P. A. M. (2003). Interdependence, interaction, and relationships. *Annual Review of Psychology, 54,* 351–375. doi: 10.1146/annurev.psych.54.101601.145059

Selten, R. (1978). The chain store paradox. *Theory and Decision, 9,* 127–159. doi: 10.1007/BF00131770

Shumway, R., & Stoffer, D. (2006). *Time series analysis and its applications: With R examples* (2nd ed.). New York: Springer.

Simon, H. A. (1956). Rational choice and the structure of the environment. *Psychological Review, 63,* 129–138. doi: 10.1037/h0042769

Simon, H. A. (1982). *Models of bounded rationality.* Cambridge, MA: MIT Press.

Starmer, C., & Sugden, R. (1991). Does the random-lottery incentive system elicit true preferences? An experimental investigation. *American Economic Review, 81,* 971–978.

Sugden, R. (1993). Thinking as a team: Towards an explanation of nonselfish behavior. *Social Philosophy and Policy, 10,* 69–89. doi: 10.1017/S0265052500004027

Sugden, R. (2005). The logic of team reasoning. In N. Gold (Ed.), *Teamwork: Multi-disciplinary perspectives* (pp. 181–199). Basingstoke: Palgrave Macmillan.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology, 46,* 35–57. doi: 10.1086/406755

Trivers, R. (2005). Reciprocal altruism: 30 years later. In C. P. van Schaik & P. M. Kappeler (Eds.), *Cooperation in primates and humans: Mechanisms and evolution* (pp. 67–83). Berlin: Springer-Verlag.

van den Bos, K., Van Lange, P. A. M., Lind, E. A., Venhoeven, L. A., Beudeker, D. A., Cramwinckel, F. M., Smulders, L., & van der Laan, J. (2011). On the benign qualities of behavioral disinhibition: Because of the prosocial nature of people, behavioral disinhibition can weaken pleasure with getting more than you deserve. *Journal of Personality and Social Psychology, 101,* 791–811. doi: 10.1037/a0023556

Van Lange, P. A .M. (1999). The pursuit of joint outcomes and equality in outcomes: An integrative model of social value orientation. *Journal of Personality and Social Psychology, 77,* 337–349. doi: 10.1037/0022-3514.77.2.337

Van Lange, P. A. M. (2000). Beyond self-interest: A set of propositions relevant to interpersonal orientations. *European Review of Social Psychology, 11*, 297–330. doi: 10.1080/14792772043000068

Van Lange, P. A. M., Bekkers, R., Schuyt, T. N. M., & Van Vugt, M. (2007). From games to giving: Social value orientation predicts donations to noble causes. *Basic and Applied Social Psychology, 29,* 375–384. doi: 10.1080/01973530701665223

Van Lange, P. A. M., W., De Bruin, E. M. N., Otten, & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology, 73,* 733–746. doi:10.1037/0022- 514.73.4.733

von Neumann, J., & Morgenstern, O. (1947). *Theory of games and economic behavior* (2nd ed.). Princeton, NJ: Princeton University Press.

Wilson, T. D., Reinhard, D. A., Westgate, E. C., Gilbert, D. T., Ellerbeck, N., Hahn, C., Brown, C. I., & Shaked, A. (2014). Just think: The challenges of the disengaged mind. *Science, 345*(6192), 75–77. doi: 10.1126/science.1250830

Yamagishi, T., Mifune, N., Li, Y., Shinada, M., Hashimoto, H., Horita, Y., Miura, A., Inukai, K., Tanida, S., Kiyonari, T., Takagishi, H., & Simunovic, D. (2013). Is behavioral pro-sociality game-specific? Pro-social preference and expectations of pro-sociality. *Organizational Behavior and Human Decision Processes, 120,* 260–271. doi: 10.1016/j.obhdp.2012.06.002

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.

Zauner, K. G. (1999). A payoff uncertainty explanation of results in experimental Centipede games. *Games and Economic Behavior, 26,* 157–185. doi: 10.1006/game.1998.0649

Footnotes

[1] Raw data are presented in the supplemental materials.

[2] Details of the time series analysis are given in the supplemental materials.

Table 1

Experiment 1: Product-Moment Correlations Between Numbers of Defections and Scores for Cooperative, Individualistic, and Competitive Social Value Orientation Under Fixed and Random Pairing in Four Games

			Social Value Orientation		
Pairing	Game	*N*	Cooperative	Individualistic	Competitive
Fixed					
	Constant difference	24	–.314	.207	.200
	Constant-sum	24	.178	–.219	.056
	Increasing difference	22	–.116	.101	.063
	Decreasing difference	22	–.262	–.119	.525
Random					
	Constant difference	24	.444	–.338	–.242
	Constant-sum	24	.178	–.210	.016
	Increasing difference	24	–.220	.154	.061
	Decreasing difference	21	–.413	.053	.605*

*Significant with a Benjamini-Hochberg false discovery rate q^* = .05.

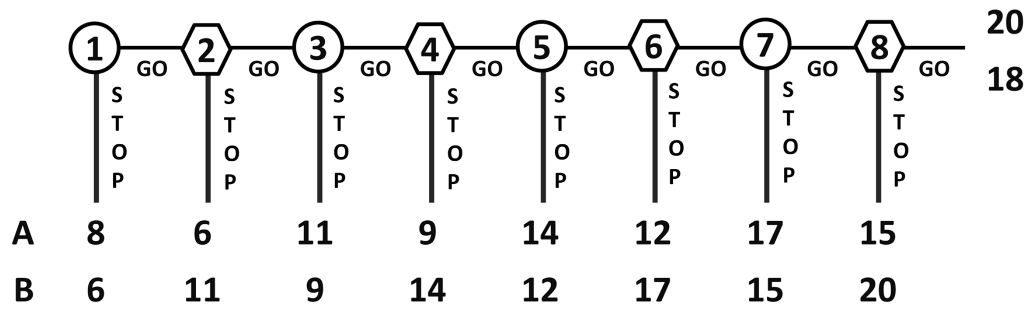
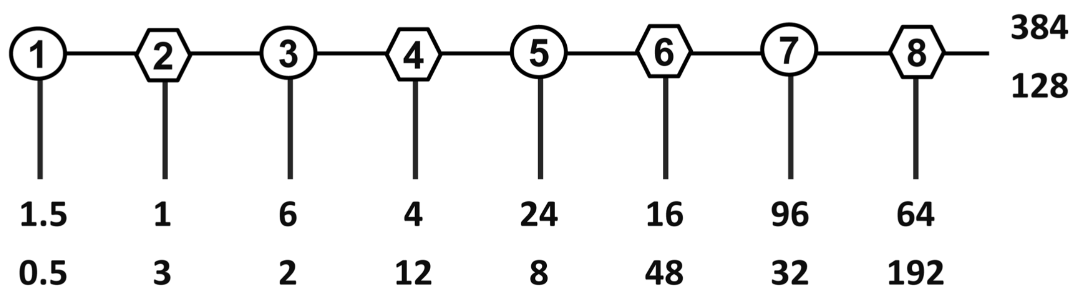


Figure 1. A linear Centipede game with $c = 2$, $b = 5$.

a.



b.

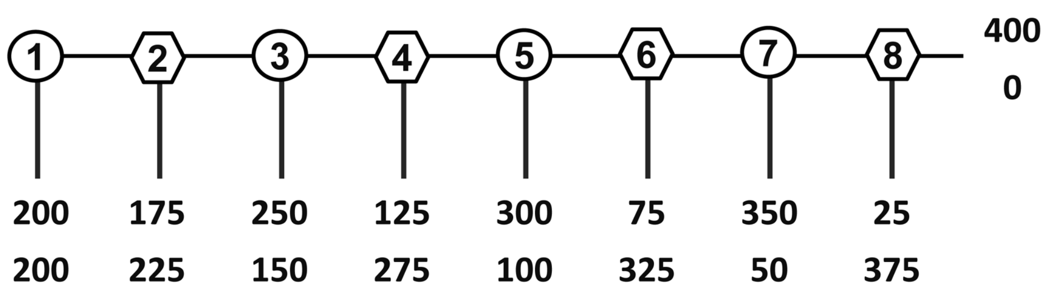


Figure 2. (a) An exponential Centipede game with $c = 1/3x$, $b = 6y$. (b) A constant-sum Centipede game with $b = c = x - y + 25$ and $x + y = 400$.

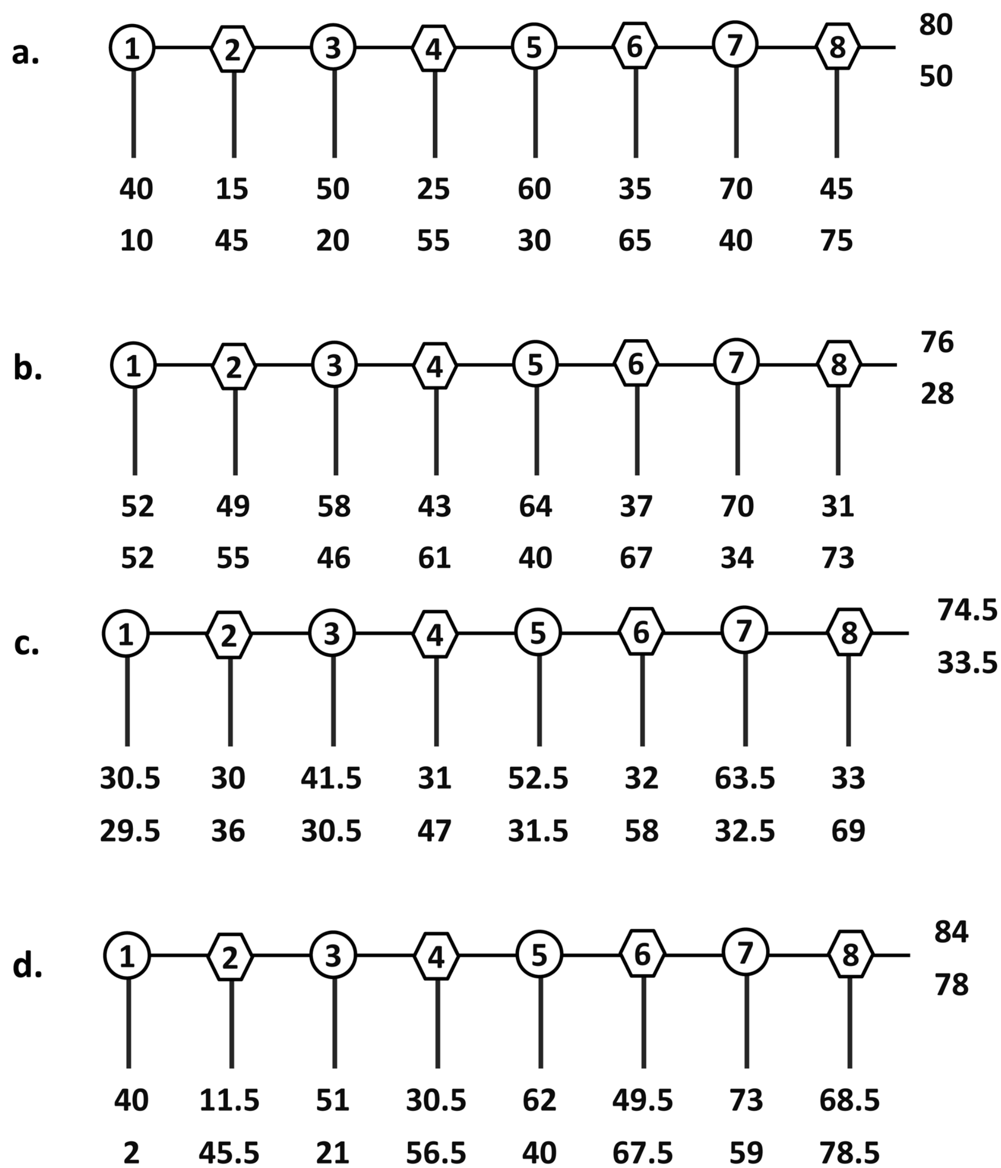


Figure 3. Linear Centipede games used in Experiment 1: (a) constant payoff-difference, (b) constant-sum, (c) increasing payoff-difference, (d) decreasing payoff-difference.

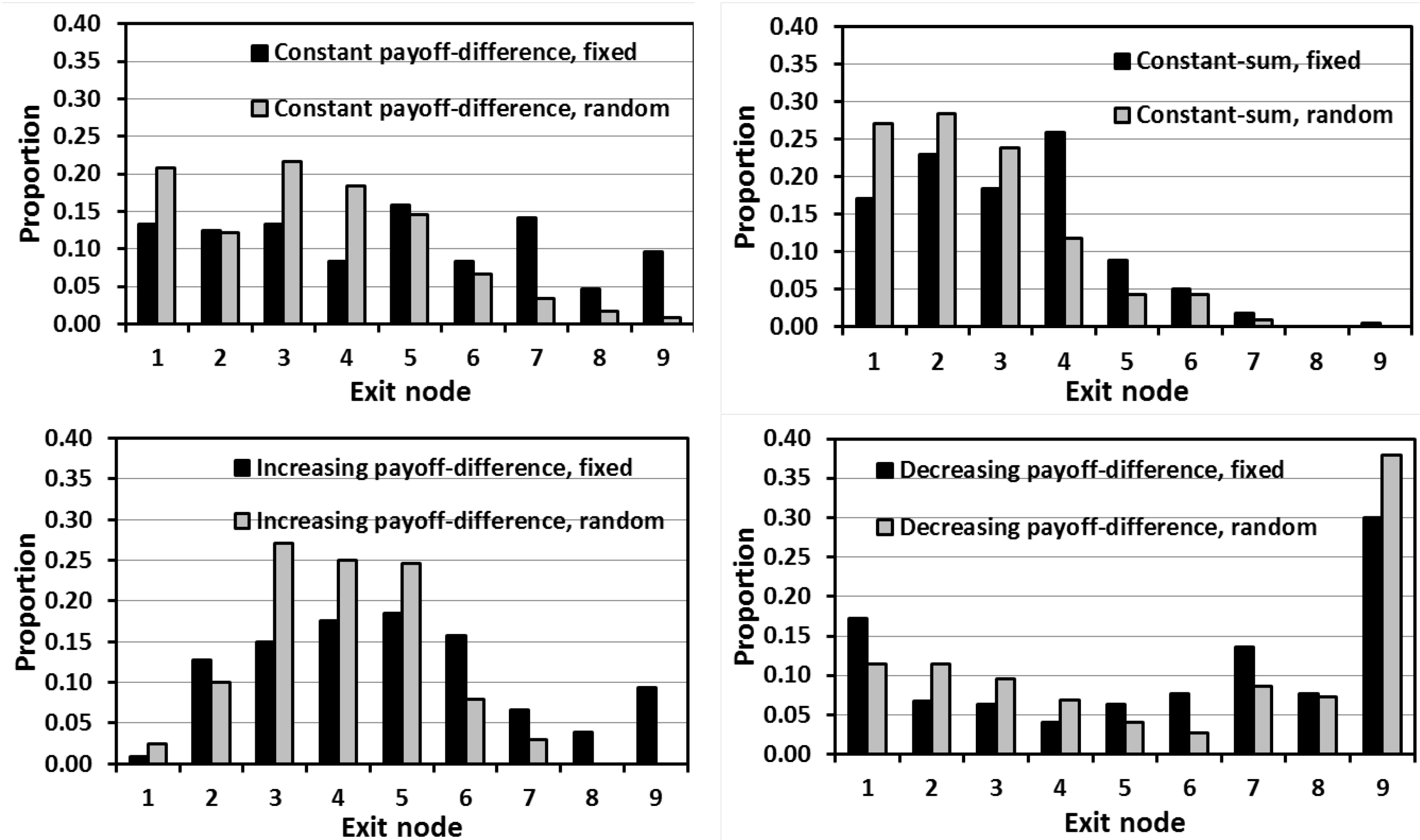


Figure 4. Experiment 1: Proportion of games ending at each terminal node under fixed or random player pairing in four experimental Centipede games: constant payoff-difference, constant-sum, increasing payoff-difference, and decreasing payoff-difference.

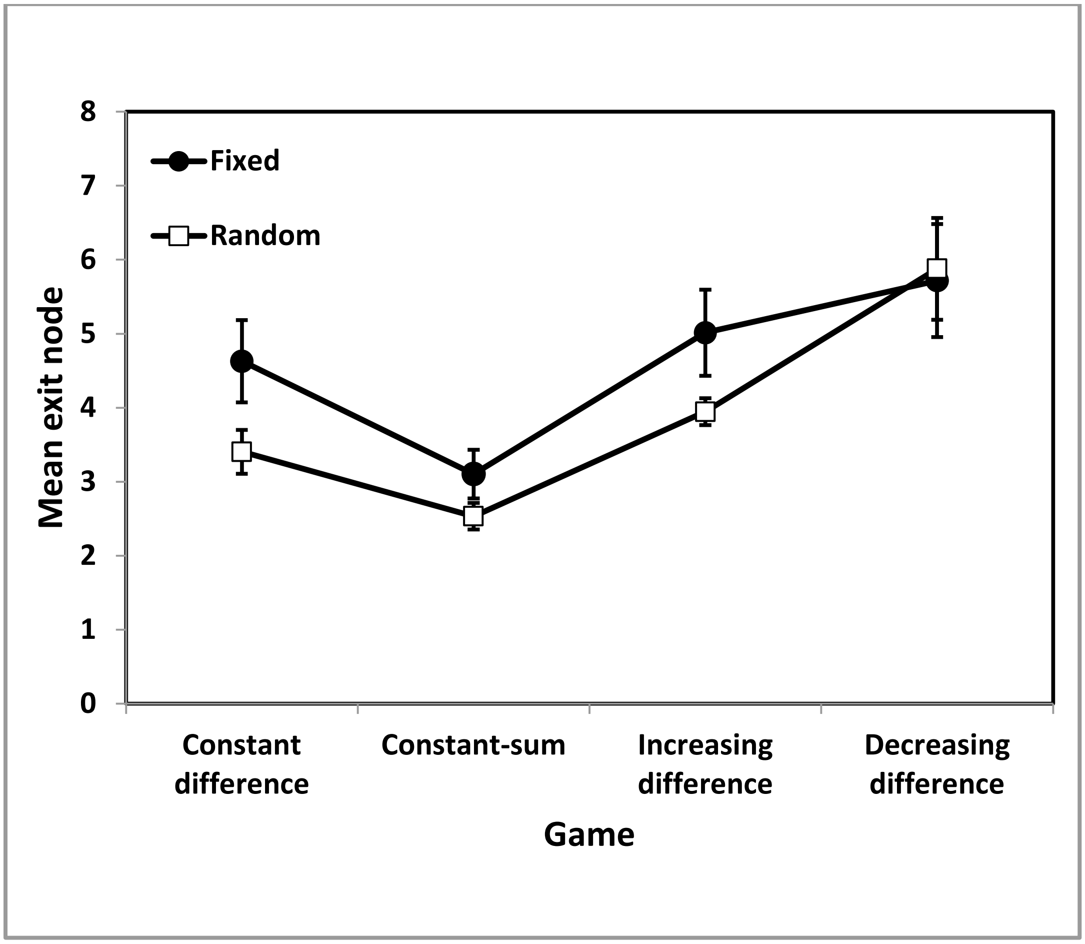


Figure 5. Experiment 1: Mean exit nodes for fixed pairing and random pairing in four Centipede games: (a) constant payoff-difference, (b) constant-sum, (c) increasing payoff-difference, (d) decreasing payoff-difference. Error bars represent standard errors of the means.

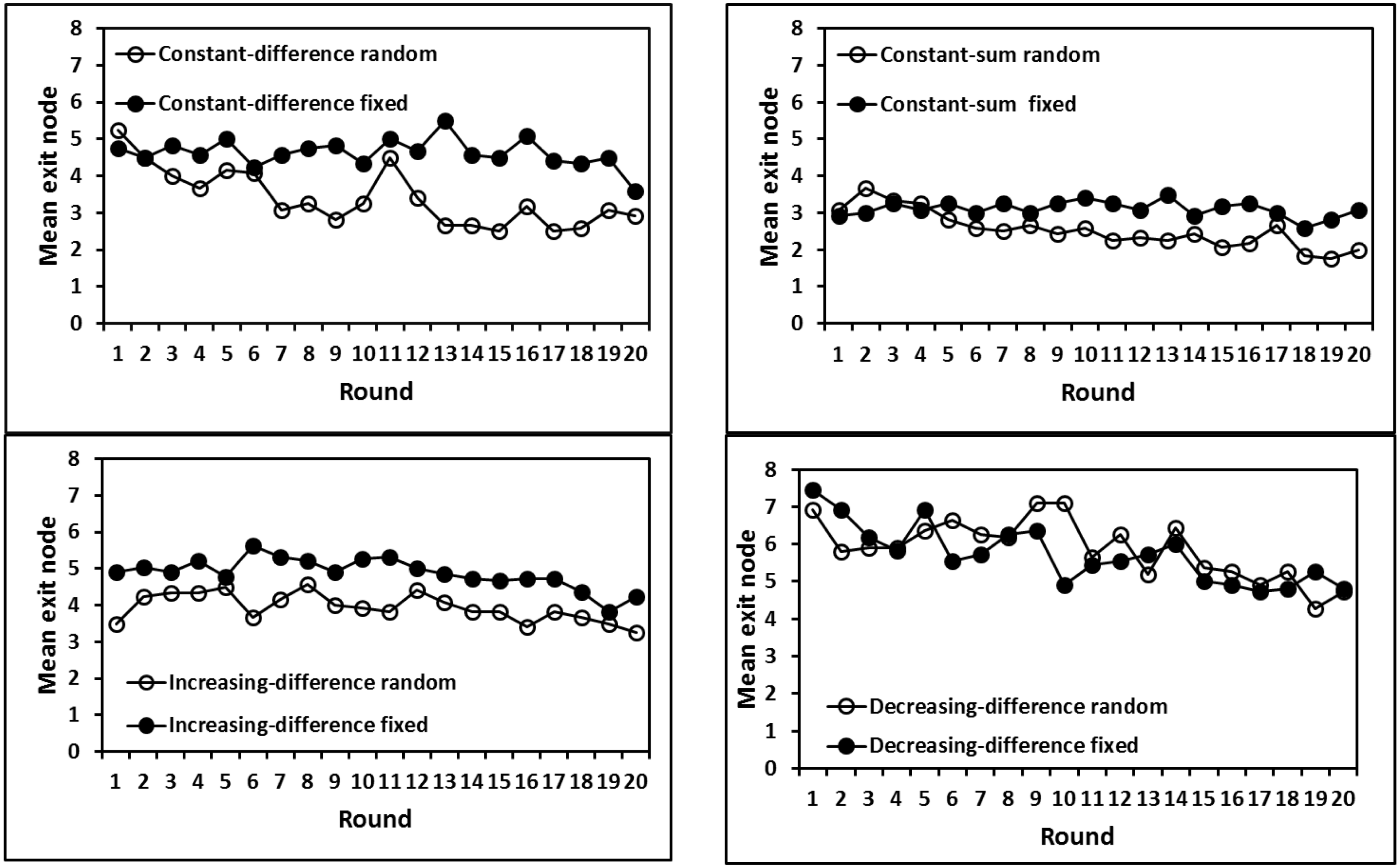


Figure 6. Experiment 1: Sequence plots of mean exit nodes under fixed pairing and random pairing in the four linear Centipede games: constant payoff-difference, constant-sum, increasing payoff-difference, and decreasing payoff-difference.

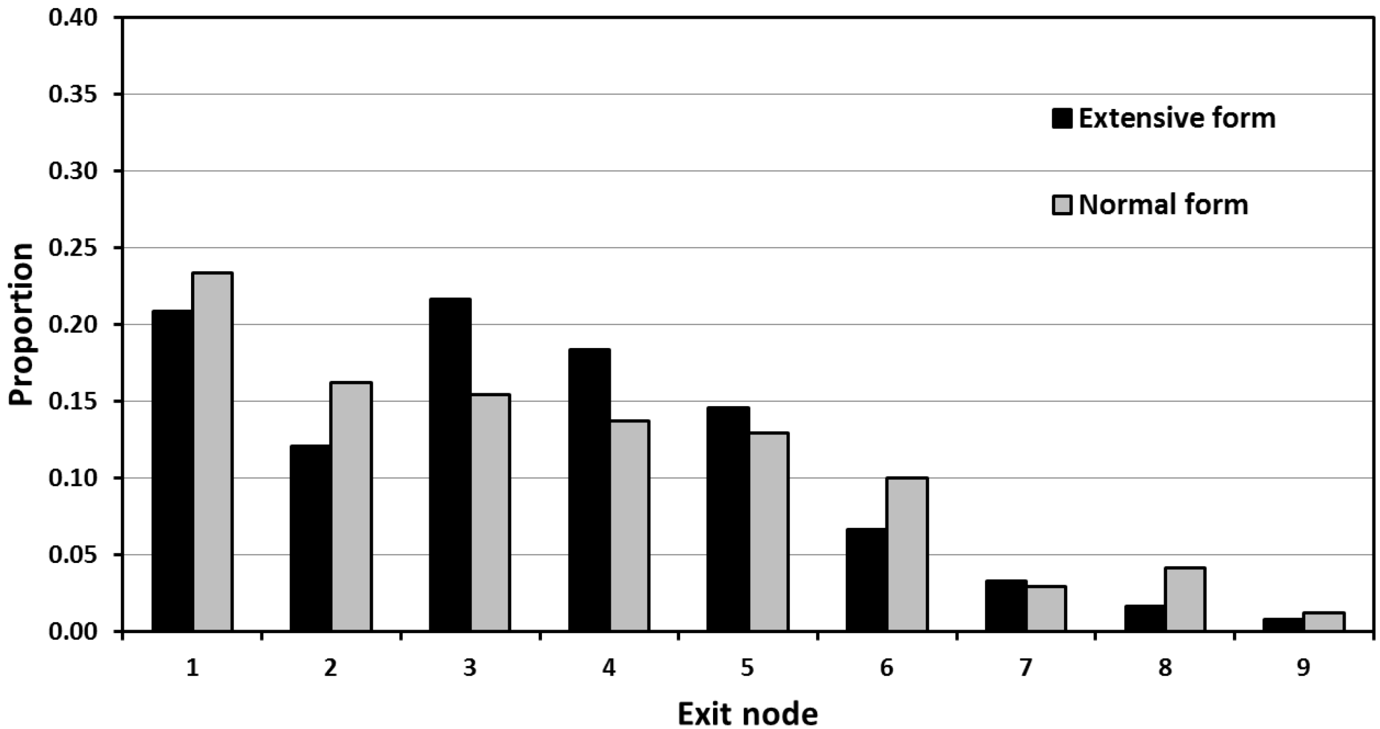


Figure 7. Experiment 2: Proportion of games ending at each terminal node in extensive form and normal form constant payoff-difference games with random player pairing.

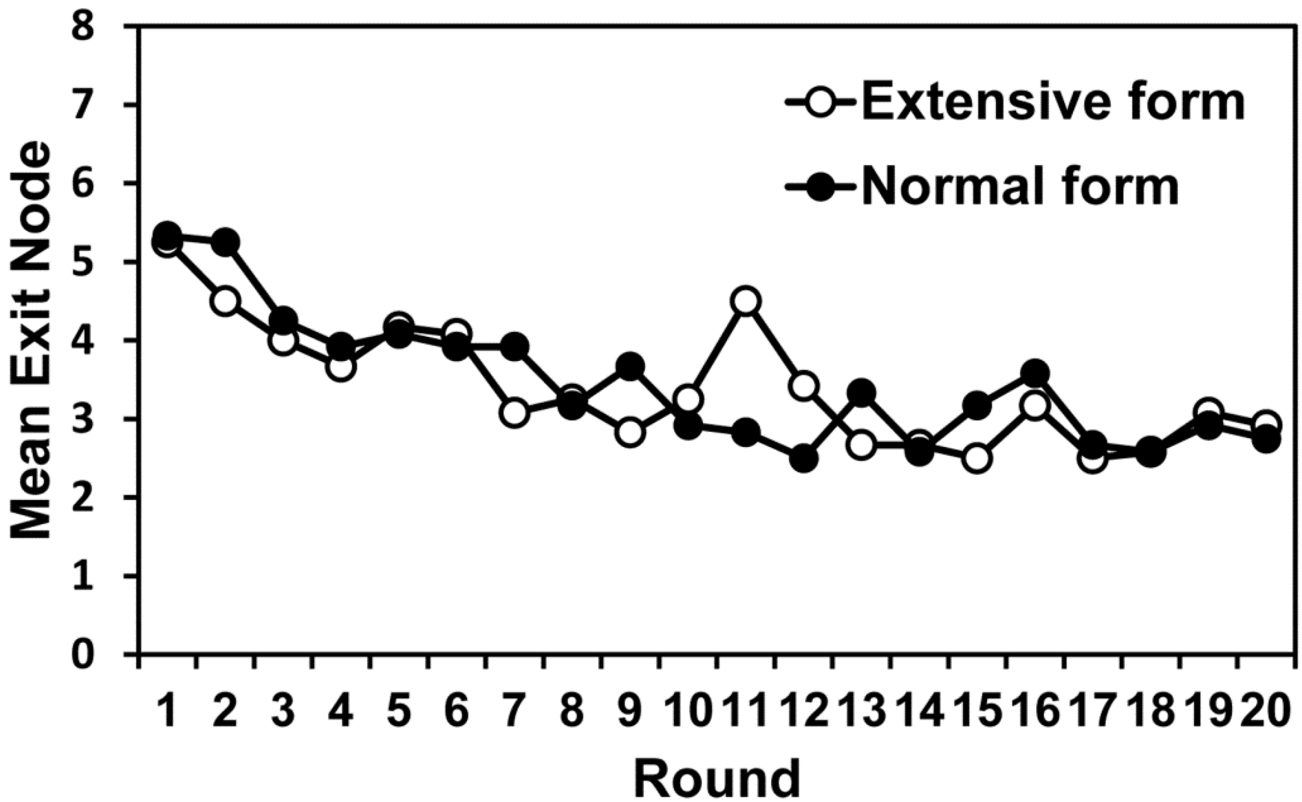


Figure 8. Experiment 2: Sequence plots of mean exit nodes for a constant payoff-difference game played in normal form and extensive form.