

Power asymmetry destabilizes reciprocal cooperation in social dilemmas

Authors

Marco Colnaghi^{1*}, Fernando P. Santos^{2†}, Paul A. M. Van Lange^{1†}, Daniel Balliet,^{1*}

¹Department of Experimental and Applied Psychology, Institute for Brain and Behaviour Amsterdam (IBBA), Vrije Universiteit Amsterdam, Amsterdam 1081BT, The Netherlands

²Informatics Institute, University of Amsterdam, Amsterdam, The Netherlands

[†]These authors contributed equally to this work

*Corresponding authors. m.colnaghi@vu.nl (MC), d.p.balliet@vu.nl (DB)

Acknowledgments

This was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement No. 864519), awarded to Daniel Balliet.

Author contributions: Conceptualization: MC, DB, FPS, PAMVL; Methodology: MC, DB, FPS, PAMVL; Software: MC; Formal Analysis: MC; Writing—original draft: MC; Writing—review & editing: MC, DB, FPS, PAMVL; Visualization: MC; Funding acquisition: DB.

Competing interests: All other authors declare they have no competing interests.

Code availability: The code used to generate the results of this study has been deposited and is available on Zenodo (<https://zenodo.org/doi/10.5281/zenodo.10687584>)

Abstract

Direct reciprocity has been long identified as a mechanism to enhance cooperation in social dilemmas. While most research on reciprocal cooperation has focused on symmetrical interactions, real world interactions often involve differences in power. Verbal theories have either claimed that power differences enhance or destabilize cooperation, indicating the need for a comprehensive theoretical model of how power asymmetries affect direct reciprocity. Here, we investigate the relationship between power and cooperation in two frequently studied social dilemmas, the prisoner's dilemma (PD) and the snowdrift game (SD). Combining evolutionary game theory and agent-based models, we demonstrate that power asymmetries are detrimental to the evolution of cooperation. Strategies that are contingent on

power within an interaction provide a selective advantage in the iterated SD, but not in the iterated PD. In both games, the rate of cooperation declines as power asymmetry increases, indicating that a more egalitarian distribution of the benefits of cooperation is the prerequisite for reciprocal cooperation to evolve and be maintained.

Keywords: Cooperation, Power, Evolutionary Game Theory, Asymmetric Games, Prisoner's Dilemma, Snowdrift Game

MAIN TEXT

Background

Social dilemmas, where individuals can increase their payoffs at the expense of collective welfare, present an important challenge to cooperation (1–4). When social interactions between individuals are repeated, direct reciprocity can provide a strong incentive towards cooperative behaviour (5–10). If the continuation probability (i.e., the likelihood of interacting again with the same individual) exceeds a critical threshold, reciprocal strategies such as Tit-For-Tat (TFT: cooperate on the first encounter, then cooperate only if one's social partner cooperated on the previous round, and defect otherwise) can outcompete defection, paving the way for cooperation to spread in a population of free-riders (7,11).

Most research on direct reciprocity in dyadic interactions has focused on symmetric games (8–13). Yet, many real-world interactions are characterized by some form of power asymmetry, where two (or more) persons differ in how strongly they can influence each another's outcomes (14–16). Humans can readily infer and respond to power asymmetries across social interactions (17–21). Power differences are widespread in the animal kingdom as well (22–24), and the role of power asymmetry in animal contests has been investigated by several theoretical studies (25–27). However, it is only in recent years that researchers have begun to study the impact of power asymmetries on direct reciprocity (28–30). Here, we offer a theoretical framework for understanding why, and under what conditions, power may or may not undermine the necessary conditions for cooperation.

Power asymmetry and cooperation

Power can be broadly defined as asymmetric control over another person's outcomes (20,31,32). Experimental and theoretical studies on power asymmetry typically focus on one (or a few) of a number of factors that can lead to asymmetric control, such as differences in

payoffs (33–36), endowments (29), effectiveness of punishment (37,38), or the ability to choose one’s social partner (39). Empirical studies that consider differences in payoffs indicate that asymmetry can pose a barrier to the emergence of cooperation in social dilemmas (33–35,39). Previous theoretical research also suggests that differences in power might destabilize cooperation in social dilemmas (28) and indicates that inequality can undermine cooperation in public goods games (29). At the same time, the effects of power differences are not unequivocal or universal; power asymmetries do not always undermine (or promote) cooperation (37,40) and power can yield different effects on cooperation in different countries (41). In fact, power hierarchies have even been considered functionally adaptive by promoting cooperation within a group (42,43). Here, we operationalize power asymmetry in a social interaction as the ability to provide higher or lower benefits to a partner in a simultaneous donation game (20,44,45), and in doing so we aim to clarify the relationship between power asymmetry and cooperation.

Inference of power asymmetry

In a social ecology with power differences across interactions, there could exist adaptive benefits to inferring power differences and conditioning behavioural strategies of cooperation on power. According to functional interdependence theory (FIT), power is one of several dimensions of interdependence that humans can infer to decide when and with whom to cooperate (46,47). FIT posits that the ability to infer and respond to power asymmetries (and other dimensions of interdependence, such as correspondence or conflict of interests) is an adaptation that provided a selective advantage in our ancestral environment (46). Indeed, humans can readily detect, and respond to, differences in power (17–21), and people employ a wide range of visual and auditory cues to infer differences in power (48–50). Children as young as five can accurately use nonverbal behaviour to discriminate between high- and low-power individuals (51), and perceived power differences can, in turn, induce major changes in one’s affective and cognitive state (21,52,53). In the context of symmetric interactions (i.e., without any differences in power), it has been shown that heterogeneous social environments can select for the ability to infer interdependence (54). In the more general case of asymmetric interactions, however, little is known about the features of the social environments that promote adaptations to infer power differences.

In the present study, we analyse the interplay between power asymmetry (in the form of unequal resource allocation) and reciprocal cooperation in iterated 2-person games. We aim to clarify how power affects reciprocal cooperation, and under what conditions can the ability to

infer power differences provide a selective advantage. We start by studying reciprocal cooperation in the iterated, asymmetric version of the simultaneous donation game, a specific form of the Prisoner's Dilemma (PD). We derive a simple formula, which expresses the minimum continuation probability necessary to stabilize cooperation as a function of power asymmetry. We then consider strategies that are conditional on power differences within interactions, and evaluate whether the ability to adapt one's behaviour in response to power asymmetry is evolutionarily stable and provides a selective advantage in the iterated PD. We subsequently shift our attention to the asymmetric version of the iterated Snowdrift game (SD; also referred to as Chicken game), another frequently studied social dilemma (9,55). While the dynamics of the iterated PD become similar to a Stag Hunt (SH) (56) if the continuation probability is high enough, the dynamics of the iterated SD between reciprocal cooperators resemble a maximizing difference/harmony game (see Methods). Therefore, focusing on the iterated SD allows us to extend our analysis to all four "archetypal" games that people most frequently use to describe social interactions (57) and are most frequently studied in the literature on social dilemmas (54,58,59). Additionally, in the SD, individuals can increase their payoffs by doing the opposite of what their social partner does; in line with previous studies, we refer to this behaviour as "anti-coordination" (see, for example, ref. (60)). Therefore, the SD allows us to study the evolution of cooperation in a situation where differences in power can help solve collective action problems (61–63).

Results

A simple rule for the evolution of reciprocal cooperation in an asymmetric PD

First, consider the simple case of two individuals, who differ in their level of power, interacting with each other through the asymmetric simultaneous donation game. In the symmetric version of this game, each individual can incur a cost c to provide a benefit b to their social partner (64). We consider an asymmetric version of this social dilemma, where high-power individuals confer a higher benefit, $(1 + \alpha)b$, and low-power individuals a lower one, $(1 - \alpha)b$, with $0 \leq \alpha < 1$. By introducing α and assuming this distribution of payoffs, we guarantee that power asymmetries do not affect the overall benefits of cooperation. We assume the cost of cooperation to be the same for high- and low-power individuals. This leads to the following payoff matrix:

High-power

		C	D	
<i>Low-power</i>	C	$\left(\begin{array}{cc} (1 + \alpha)b - c, (1 - \alpha)b - c & -c, (1 - \alpha)b \\ (1 + \alpha)b, -c & 0, 0 \end{array} \right) \quad (1)$		
	D			

Using the framework of interdependence theory (46,47), it is possible to show that, for the parameters used throughout the study, the degree of asymmetric dependence (a method of assessing power asymmetry) increases monotonically with α (Methods; Figure S1). In other words, the greater the asymmetry (α), the more a low-power individual's outcome is influenced by their partner's choice to cooperate or not. For simplicity, throughout the paper we refer to α as power asymmetry. This definition of power is consistent with theoretical work that suggests that the ability to allocate greater rewards is an expression of power (20,44,45).

In order to study the impact of direct reciprocity, we assume that there is a continuation probability w of repeated interaction with the same individual, and individuals can either choose reciprocal cooperation (TFT) or always defect (AllD). We limit our analysis to these two strategies as we are interested in the necessary conditions that promote the evolution of reciprocal cooperation in a population of defectors; if TFT cannot displace AllD, no other strategy can (7). Moreover, when cooperation costs are high, TFT remains a key strategy to sustain cooperation even when stochastic and longer-memory strategies are considered (65). When only these two strategies are considered, it is possible to prove (see Methods) that TFT is an evolutionarily stable strategy (ESS) if, and only if, the continuation probability exceeds the threshold:

$$w^* = \left(\frac{1}{1 - \alpha} \right) \frac{c}{b} \quad (2)$$

For $\alpha = 0$, we retrieve the well-known condition for the evolution of cooperation in the symmetrical PD, $w > c/b$ (7). When $\alpha > 0$, any increase in power asymmetry leads to a higher threshold continuation probability, making it harder for cooperation to evolve (Figure 1). As α approaches the value of 1, the minimum continuation probability necessary for cooperation to be stable diverges to infinity. This simple formula illustrates the negative impact that power asymmetry has on cooperation: even if w is high enough to sustain cooperation in the symmetric case, increasing power inequality will eventually lead to the breakdown of cooperation.

Evolutionary stability and fixation probability

Next, we consider the more complex case of a population where individuals have an equal probability of finding themselves in a position of high or low power. This assumption reflects the empirical observation that most variation in perceived power is due to changes in situations, rather than stable individual traits (21). Individuals can therefore find themselves in three types of possible interactions: symmetric, low-power (when a low-power individual interacts with a high-power one), and high-power (when a high-power individual interacts with a low-power one). To study whether the ability to infer differences in power can provide a selective advantage in the PD, we consider eight possible strategies: two “power-independent” strategies, where individuals always play TFT or AllD in every situation regardless of power asymmetries, and six strategies conditional on power, where individuals choose whether to play TFT or AllD depending on the specific type of interaction (symmetric or with a higher/lower-power individual). Each of the eight strategies can then be identified as a triplet (X, Y, Z) where an individual plays X in symmetric interactions, Y when in low power and interacting with a higher-power individual, and Z when in high power and interacting with lower-power individual. From simplicity, we use the shorthand “C” to refer to reciprocal cooperation (TFT), and “D” as a shorthand for AllD. For example, the strategy (C, C, D) plays TFT in symmetric interactions and when interacting with higher-power individuals, and AllD when interacting with individuals with a lower power status.

Using a standard evolutionary game-theoretical approach (66), we evaluate under what conditions each of these strategies is evolutionarily stable (Figure 2). In addition, we use a small-mutation approximation (67) to evaluate whether each of the strategies that are evolutionarily stable in an infinite population would evolve in a finite population subject to stochastic perturbations (Figure 3). The ESS analysis shows that, for a given level of power asymmetry, a cooperative power-independent strategy (C, C, C) is evolutionarily stable, provided that the continuation probability is high enough (Figure 2). As power asymmetry increases, the minimum threshold that makes reciprocal cooperation advantageous increases as well, making cooperation less stable. As a single cooperative mutant can never invade an infinite population of defectors, (D, D, D) is an ESS in the whole parameter space (Figure 2). In addition to these two power-independent strategies, two mixed strategies, (D, C, C) and (C, D, D) , are evolutionarily stable in two regions of the parameter space. However, a finite-population size analysis reveals that these two strategies almost never outperform power-independent ones (Figure 3). (C, D, D) outperforms other strategies only for a limited range of

continuation probability w and in large populations ($N = 500, 1000$; Fig. S2). (D, C, C) is never advantageous in a finite population. In principle, these strategies are stable once they spread to fixation in an infinite population; in practice, however, natural selection does not promote their spread (if not in a very narrow range of ecological conditions), making the fixation of power-independent strategies more likely. When repeated interactions are infrequent, natural selection favours defection in all situations (D, D, D) (Figure 3). As the continuation probability increases, so does the equilibrium frequency of (C, C, C) , eventually outcompeting (D, D, D) : once cooperation becomes advantageous, it is favorable to cooperate in all situations, regardless of whether the interaction is symmetric or asymmetric (Figure 3). To conclude, the ability to infer power and modify one's behaviour accordingly is unlikely to provide a selective advantage in the iterated asymmetric Prisoner's Dilemma.

Asymmetric Snowdrift game

The simultaneous donation game described above can be modified to assume the form of a Snowdrift (SD) game, a slightly more benign social dilemma where cooperators and defectors can stably coexist in a population (9,55). This game represents a situation where two individuals must work together to achieve a certain outcome, such as building a shelter or freeing a road from a snowdrift, which benefits them both. Each individual would be better off if their partner would bear the whole cost of this enterprise, but incurring the cost of cooperation alone is preferable to the situation where neither player contributes to the common good.

Let $2b$ be the total payoff obtained through mutual cooperation. Suppose that the total cost of the endeavor, $2c$, can be either split between the two individuals or borne by one individual alone. As in the previous case, the reward can be either divided symmetrically when two individuals of equal power meet, or asymmetrically, if there is a power difference between the two players. In the former case, both players receive the same reward b . In the latter, the high-power player shares with the other a greater proportion of the reward, $(1 + \alpha)b$, and takes a smaller proportion, $(1 - \alpha)b$, for themselves. This specific form of SD game allows us to compare the results directly with the asymmetric PD, as the resulting payoff matrix has the same reward for mutual cooperation:

$$\begin{array}{cc}
 & \text{High-power} \\
 & \begin{array}{cc}
 & \text{C} & \text{D} \\
 \text{Low-power} & \begin{array}{cc}
 \text{C} & \left((1 + \alpha)b - c, (1 - \alpha)b - c \right) & (1 + \alpha)b - 2c, (1 - \alpha)b \\
 \text{D} & (1 + \alpha)b, (1 - \alpha)b - 2c & 0, 0
 \end{array}
 \end{array}
 \end{array} \quad (3)$$

Again, we assume a continuation probability w of repeated interaction with the same partner, and consider two different behaviors: reciprocal cooperation (TFT) or always defect (AllD). In the simple case of a high-power individual interacting with a low-power one, we retrieve the same simple rule expressed by Equation (2). In the more complex case of a population where individuals can find themselves in a high- or low-power position with equal probability, we consider the two power-independent and the six strategies conditional on power introduced in the previous section (see Methods). Again, we evaluate which strategies are ESSs, and apply a small-mutation approximation to analyse which of the ESSs are favoured by natural selection in a finite population.

A standard evolutionary game-theoretical analysis (66) reveals that (as in the case of the PD) a fixed strategy of reciprocal cooperation (C, C, C) becomes evolutionarily stable once the continuation probability is high enough and power asymmetry is not too extreme (Figure 4). As power asymmetry increases, (C, C, C) becomes unstable and is replaced by a strategy conditional on power differences, (C, C, D) , which plays TFT in symmetric interactions and when interacting with higher-power individuals, and defects when interacting with lower-power individuals (Figure 4). Performing a small-mutation, finite population size analysis, we find that this strategy outcompetes both power-independent strategies, (C, C, C) and (D, D, D) , when the continuation probability is low (Figure 5). As w increases, (C, C, C) becomes more favorable, and it eventually outcompetes (C, C, D) (Figure 5). These conclusions are not affected by population size (Figure S3). Thus, if the continuation probability does not exceed this threshold, the ability to infer power and modify one's behavior accordingly provides a selective advantage in the iterated asymmetric SD.

Changes in cooperation frequency

Finally, we consider how the introduction of strategies conditional on power changes the average level of cooperation in a finite population under a small-mutation approximation, calculated as the equilibrium frequency of a strategy times the fraction of interactions where that strategy will play TFT (Figure 6). These results confirm our claim that cooperation

declines with power asymmetry (Figure 6). This effect becomes more pronounced when the continuation probability is higher, and the decline in average cooperation increases with w . Thus, higher levels of power asymmetry cause a decline in the frequency of cooperative interactions.

Discussion

In the symmetric prisoner's dilemma (PD), direct reciprocity is evolutionarily stable if, and only if, the continuation probability (i.e., the likelihood of repeated encounters with the same individual) exceeds the cost-to-benefit ratio (7). We show that power asymmetry, modeled as the ability to allocate smaller or greater rewards in a simultaneous donation game, destabilizes reciprocal cooperation, raising the minimum continuation probability needed for cooperation to evolve (Equation 2, Figure 1). An intuitive explanation for this phenomenon is that the cost-to-benefit ratio of cooperation for high-power individuals (who can allocate greater rewards, and receive smaller ones) increases with power asymmetry: if the level of asymmetry exceeds a critical threshold, the reward for mutual cooperation is so small that exploiting one's social partner once becomes more advantageous than sustained reciprocal cooperation. Therefore, it is in the interests of high-power individuals to defect, even when repeated interactions afford the possibility of reciprocal cooperation.

The ability to infer power differences across different interactions (symmetric, low-power, and high-power) does not provide any selective advantages in the PD. In this social dilemma, power-independent strategies always outcompete strategies that are contingent on power asymmetries, except for a very narrow range of continuation probability (Figure 3). By contrast, in the SD, a strategy that cooperates in symmetric and low-power interactions and defects in high-power interactions, (C, C, D) , is evolutionarily stable in infinite populations (Figure 4) and favored by natural selection in finite populations (Figure 5) for low to intermediate values of the continuation probability. Among the strategies that are conditional on power, (C, C, D) is the one that achieves the highest payoffs: when (C, C, D) interacts with itself in asymmetric interactions, the low-power individual cooperates, and the high-power individual therefore receives the highest payoff available to them. This strategy can outcompete (C, C, C) by exploiting it in low-power interactions, and (D, D, D) by cooperating with itself in symmetric interactions. Perhaps less obviously, (C, C, D) can also outcompete other strategies conditional on power differences, such as (C, D, C) . For example, a rare (C, C, D) mutant can invade a (C, D, C) population, but not vice versa, as the payoff of a low-

power individual cooperating with a high-power individual is higher than that of a high-power individual cooperating with a low-power one.

This result suggests that the ability to infer differences in power provides a selective advantage in environments characterized by social dilemmas where anti-coordination is beneficial (i.e., when individuals can increase their payoffs by doing the opposite of what their partner does), such as the SD. In these games, the total payoffs are higher if individuals make opposite choices, especially because unilateral defection yields the best possible outcome whereas mutual defection yields the worst possible outcome (46,47). By contrast, when the continuation probability is high, the iterated PD takes the form of a Stag Hunt (56); in this type of game, the total payoffs are higher if individuals make matching choices, especially because mutual cooperation yields the best possible outcome whereas unilateral cooperation yields the worst possible outcome.

It has been suggested that the emergence of power asymmetries, such as in the form of leadership and followership, provided a selective advantage in ancestral human societies, as it helped to solve collective action problems (61–63). Repeated social interactions, in a context where a mixed group of leaders and followers performs better than a uniform group, give rise to dynamics typical of the SD: if everyone strives to be a leader, an individual can increase their own (and collective) payoffs by being a follower; in a population of followers, becoming a leader will also lead to higher payoffs. Our model indicates that, under these conditions, the ability to infer differences in power helps maximize joint payoffs by facilitating anti-coordination, and the evolution of decision rules based on power can therefore be promoted by natural selection. We do not observe this in the iterated PD, where individuals can increase their payoffs by matching their partner's choice. In these games, strategies that discriminate based on power perform poorly against themselves, because they involve doing the opposite of what one's partner does when interactions are asymmetric.

Power asymmetries are often addressed by cues that are shared between people or other animals. Social status is an example of cue that can be used to coordinate in social interactions. Prior research shows that people tend to defer to the preferences of higher-status individuals, even when status is arbitrarily assigned (68–70). This has been found to be particularly advantageous in asymmetric games, such as the battle of the sexes, where coordination would otherwise be difficult to achieve (70). While social status and power are distinct constructs, they are closely associated, and cues of status can be used to infer power (19,20).

Empirical studies on asymmetric games support the claim that asymmetry destabilizes cooperation (33–36,39). In asymmetric PDs, cooperation rates are significantly lower than in symmetric ones (34,35). Moreover, information about the payoff matrix reduces cooperation rate in asymmetric games (36), while increasing it in symmetric ones (36,71). Experimental studies also indicate that the temptation to defect is stronger for the players who have less to gain from cooperation (corresponding to the high-power players in our model) (34–36). By contrast, in repeated asymmetric games, players who receive higher payoffs (corresponding to low-power individuals in our model) are more likely to initiate cooperation and less likely to defect in response to their partner's defection (34). These results indicate that, in agreement with our theoretical predictions, individuals who have less to gain from cooperation aim to minimize payoff disparity through frequent defection, thereby undermining the emergence of stable cooperation. While this behaviour is a suboptimal strategy in the PD (Figure 3), it can provide a selective advantage in the asymmetric SD (Figures 4 and 5).

Our theoretical results suggest further experimental studies to investigate whether the tendency to defect when in the high-power state is stronger in asymmetric SD than PD games, which would support the hypothesis that this behaviour confers a selective advantage in games where anti-coordination is on average beneficial. While the empirical evidence discussed above indicates that payoff differences hinder cooperation, there is no universal consensus on the effects of power, and different sources of power asymmetry may yield distinct outcomes (37,40,41). For example, differences in the effectiveness of punishment promote cooperation in a modified version of the PD, where players can contribute a variable investment to the common good (37), but not in a standard PD, where players only have a binary choice to cooperate or defect (38). The asymmetric ability to punish is also more effective in promoting cooperation when used to encourage joint cooperation, rather than to exploit one's partner, in PD experiments (41). Different forms of asymmetry can also act synergistically with one another, as shown by a recent theoretical analysis of asymmetric public goods games (29). This study shows that extreme power asymmetry, in the form of an unequal distributions of endowments, prevents the emergence of reciprocal cooperation; but when the rewards of cooperation are also asymmetric, unequal endowments actually promote, rather than hinder, cooperative behaviour (29). Taken together, these studies suggest future theoretical and empirical work to further examine how different bases of power impact cooperation across different games.

Our results are based on the assumption that individuals have an equal likelihood of finding themselves in a high- or low-power position. This assumption reflects the empirical observation that variations in people's perceptions of power are mainly due to changes in situations, rather than their stable traits (21). In addition, we used variations of the simultaneous donation game as a framework to study the emergence of reciprocal cooperation. Including in our analysis both PD and SD games and varying the continuation probability allow us to explore a space of the four "archetypal" games most frequently studied in the literature on social dilemmas (54,57–59); yet, this is but a fraction of the 8-dimensional parameter space of all asymmetric, dyadic games. Further studies are needed to determine to what extent our conclusions can be generalized across all possible games.

Another important question concerns the exploration of alternative strategies of reciprocal cooperation. We only focused on TFT and AllD: this is because under the assumptions of our model (no noise and no costs associated with TFT compared to other strategies), if TFT cannot outcompete AllD, no other cooperative strategy can (7). While TFT may pave the way for cooperation to evolve, it is not necessarily the best strategy to maintain cooperation once it has been established (11,12). In fact, TFT can be invaded by AllC through drift, making it possible for AllD to eventually take over the population (11); on average, however, these evolutionary cycles tend to favour TFT (72). As the main focus of this work is to establish the necessary conditions for cooperation to emerge, and TFT invading AllD is the minimum requirement for cooperation to evolve (7), we did not consider other strategies than TFT and AllD. Future investigations on cooperation in asymmetric games could explore a broader space of strategies, such as zero-determinant strategies (73,74), win-stay lose-shift (11,12), or forgiving TFT(75), and study the impact of alternative update rules (76).

It is often said that "power corrupts", expressing the common wisdom that individuals in high-power positions can succumb to the temptation to exploit their subordinates (77). Our results indicate that, when power is operationalized as the ability to provide higher rewards, power asymmetry makes it harder for reciprocal cooperation to evolve: as low-power individuals cannot provide enough benefits to their social partners to even out the costs of cooperation, high-power individuals have a stronger incentive to defect. However, conditioning one's decision to cooperate on the level of power is not always an optimal strategy. In the iterated PD, it is more beneficial to adopt a power-independent strategy of either reciprocal cooperation (TFT) or defection (AllD), regardless of power status (Figures 2 and 3). In the iterated SD, on the other hand, power offers the means to maximize collective

payoffs through anti-coordination: low-power individuals benefit from “carrying” the cost of cooperation, as they will receive the same treatment when they find themselves in a high-power position (Figures 4 and 5). Yet, regardless of whether the optimal strategy is conditional or not on the level of power, higher levels of asymmetry hinder the emergence of cooperation (Figure 1) and thus lowers cooperation rates (Figure 6). Power asymmetry undermines reciprocal cooperation (Figures 1 and 6), and interventions to enhance cooperative behaviour should aim at promoting a more egalitarian profitability of mutual cooperation.

Methods

Operationalization of power

Interdependence theory provides the means of quantifying the exact amount of power asymmetry of an interaction described by a given payoff matrix (46,47). The variance in outcomes of an interaction can be divided in three distinct components: actor control (AC), partner control (PC), and joint control (JC). These values indicate to what extent the variance in outcomes is determined, respectively, by the player’s choices, by their partner’s choices, or by the importance of coordination (that is, the difference between making the same choice as one’s social partner or doing the opposite of what they do).

Let $G_1 = \begin{pmatrix} a_1 & b_1 \\ c_1 & d_1 \end{pmatrix}$ be the payoff matrix for player 1. Then, $AC_1 = \frac{1}{2}(a_1 + b_1 - c_1 - d_1)$, $PC_1 = \frac{1}{2}(a_1 - b_1 + c_1 - d_1)$, and $JC_1 = \frac{1}{2}(a_1 - b_1 - c_1 + d_1)$. These quantities can be used to calculate the degree of mutual dependence (MD) of an interaction, which quantifies to what extent player 1’s payoff depends on player 2’s choices:

$$MD_1 = \frac{PC_1^2 + JC_1^2}{AC_1^2 + PC_1^2 + JC_1^2} \quad (4)$$

Mutual dependence varies between 0 (no dependence) and 1 (total dependence). In asymmetric interaction, this quantity can differ between the two players; for example, if $MD_1 > MD_2$, player 2’s choices affect player 1’s outcome to a greater extent than player 1’s choices affect player 2’s outcome. The degree of asymmetric dependence, the difference in mutual dependence between the two players ($MD_1 - MD_2$), indicates the asymmetric ability to control another person’s outcomes. As shown in Figure S1, this quantity increases monotonically with power asymmetry (α) as we defined it in the context of the asymmetric donation game.

Evolutionary stability of TFT

Let w be the continuation probability (i.e., the likelihood of repeated encounters between the same individuals). In the PD, when only two strategies, TFT and AllD, are considered, and the interaction is characterized by power asymmetry α , the result is a bimatrix game with the following payoff structure:

$$\begin{array}{cc}
 & \begin{array}{cc} \text{High-power} \\ \text{C} & \text{D} \end{array} \\
 \begin{array}{c} \text{Low-power} \\ \text{C} \\ \text{D} \end{array} & \left(\begin{array}{cc} \frac{(1+\alpha)b-c}{1-w}, \frac{(1-\alpha)b-c}{1-w} & -c, (1-\alpha)b \\ (1+\alpha)b, -c & 0,0 \end{array} \right) \quad (5)
 \end{array}$$

The necessary conditions for TFT to be evolutionarily stable for both players are $\frac{(1+\alpha)b-c}{1-w} > (1+\alpha)b$ and $\frac{(1-\alpha)b-c}{1-w} > (1-\alpha)b$. This leads to $w > \frac{1-c}{1+\alpha b}$ and $w > \frac{1-c}{1-\alpha b}$. As $\alpha > 0$, the first condition is always satisfied if the second is. Using a similar reasoning, it is possible to obtain the same conditions for the iterated SD (as the reward for mutual cooperation and the temptation to defect are the same in the two games, the conditions are the same). Here we focus on TFT and AllD alone as the evolutionary stability of TFT against AllD is a necessary condition for the evolution of reciprocal altruism (7).

Evolutionary stability of strategies conditional on power

We consider an infinite population where individuals can randomly find themselves in a high- or low-power position, and interact with random members of the population. Therefore, every individual engages in symmetric interactions with a frequency of 0.5, in interactions with a lower-power individual with a frequency of 0.25, and in interactions with a higher power individual with a frequency of 0.25. The total payoff is averaged over a large number of interactions. We assume the population to be well-mixed. In each interaction, depending on the level of power asymmetry, an individual can choose whether to adopt reciprocal cooperation (TFT) or always defect (AllD). This gives rise to eight possible strategies: two “power-independent” strategies, where individuals always play TFT or AllD in every situation regardless power differences, and six strategies contingent on power differences, where individuals choose whether to play TFT or AllD depending on the specific type of interaction (symmetric or with a higher/lower-power individual). We do not consider strategies that distinguish between low-power and high-power symmetric interactions as we are mainly

interested in studying the stability of strategies that depend on power asymmetries. Each of the eight strategies can then be identified as a triplet (X, Y, Z) where an individual plays X in symmetric interactions, Y when in low power and interacting with a higher-power individual, and Z when in high power and interacting with lower-power individual. From simplicity, we use the shorthand “C” to refer to reciprocal cooperation (TFT), and “D” as a shorthand for AllD. For example, the strategy (C, C, D) plays TFT in symmetric interactions and when interacting with higher-power individuals, and AllD when interacting with individuals with a lower power status. Assuming that individuals have the same probability of being in a low or high-power state, and this state can eventually change between each interaction, the total payoff of an individual adopting strategy i is $P_i = \frac{1}{2}P_i^S + \frac{1}{4}P_i^H + \frac{1}{4}P_i^L$, where P_i^S , P_i^H , and P_i^L are, respectively, the average payoff of strategy i in symmetric, high-power, and low-power interactions, which will in turn depend on the frequencies of other strategies in the population. Let P_{ji} be the average payoff of a rare mutant playing strategy j in a population where strategy i is fixed. Strategy i is evolutionarily stable if, and only if, $P_{ii} > P_{ji}$ for every other strategy j , or if $P_{ii} = P_{ji}$ and $P_{ij} > P_{jj}$ (66). We evaluated numerically which strategies are evolutionarily stable for given values of α and w .

Small-mutation approximation

We also analyse the evolutionary dynamics of the strategies defined above in a finite population using a small mutation approximation, i.e., assuming that the time between the emergence of a new mutant playing a different strategy is much greater than the time it takes for a mutant to go extinct or spread to fixation (67). We model the evolutionary dynamics of the population as a Moran process, where at each time step an individual is randomly selected proportionally to their fitness to replace another, randomly selected individual (78,79). We define the fitness of an individual adopting strategy i as $f_n(i) = e^{\beta P_n^i}$, where β denotes the strength of selection and P_n^i denotes the payoff of an individual playing strategy i in a population where there are exactly n individuals adopting strategy i (79). Let $x(i)$ be the frequency of strategy i . The probability that an individual adopting strategy i is selected for reproduction is $x(i)f_n(i) / \sum_k x(k)f_n(k)$, where the denominator is the average fitness of the population. The offspring of the individual selected for reproduction replaces another individual, selected at random independently of fitness; that is, the probability that an individual adopting strategy j is replaced is $x(j)/N$.

We model the evolution of this population as a discrete time Markov chain with transition matrix T , whose elements T_{ij} indicate the transition probability from state j (where strategy j is fixed in the population) to state i (where strategy i is fixed). The transition probability from j to i is given by the product between population size (N), mutation rate (U), and the rate of evolution ρ_{ij} , i.e., the probability of fixation of a single mutant adopting strategy i in a population of j (80). As the stationary distribution of a Markov chain does not change if all non-diagonal entries of the transition matrix are multiplied by the same factor, we can omit the terms N and U and write the transition coefficients as:

$$T_{ij} = \left(1 + \sum_{k=1}^{N-1} \prod_{n=1}^k \frac{f_n(j)}{f_n(i)} \right)^{-1} \quad (6)$$

As there are no absorbing states, this matrix represents an ergodic Markov chain with a unique stationary distribution, which can be calculated as the left eigenvector associated with the eigenvalue $\lambda = 1$ of the transition matrix.

Exact stationary distribution

In order to evaluate whether the small-mutation approximation is a reasonable assumption to describe the evolution of such population, we consider the simplified case of only three strategies, for which the stationary distribution can be calculated exactly. For the PD, we considered the two power-independent strategies and one of the evolutionarily stable strategies conditional on power, (C, D, D) . For the SD, we considered the two power-independent strategies and the only evolutionarily stable strategy conditional on power, (C, C, D) . These strategies were selected as they can be evolutionary stable for certain values of w and α in an infinite population (Figure 2 and 4) and outcompete other conditional strategies in a finite population (Figure 3). The population evolves according to a Moran process, as described in the previous sections, with the only difference that, every time an individual reproduces, it mutates to a different strategy with probability U . At each time step, an individual adopting strategy i is selected with probability $x(i)f_n(i) / \sum_k x(k)f(k)$. With probability $(1 - U)$, their offspring adopts strategy i . With probability U , the offspring adopts a different strategy (the offspring has an equal probability of adopting any other strategy). As previously, the offspring replaces a randomly chosen individual, independently of their fitness.

Let (k, p, q) denote a state with k individuals adopting (C, C, C) , p individuals adopting (D, D, D) and q individuals adopting (C, C, D) , and let $x = k/N$, $y = p/N$, and $z = q/N$ be the frequencies of these strategies in the population. The transition probability from (k, p, q) to $(k + 1, p - 1, q)$ is:

$$T_{k,p,q \rightarrow k+1,p-1,q} = xy \left(1 - \frac{3}{2}U\right) f^{CCC}(k, p, q) + \frac{U}{2}y \quad (7)$$

Where $f^{CCC}(k, p, q)$ designs the relative fitness of (C, C, C) in state (k, p, q) . Analogous formulae describe the transitions from (k, p, q) to five other possible adjacent states. The transition probabilities are calculated numerically, and the transition matrix is then used to calculate the stationary distribution. As shown in Figures S4 and S5, up to a mutation rate of $U = 0.01$, the system spends a negligible time in the intermediate states, the pure states (where one strategy has reached fixation in the population) being the ones with the highest frequency (Figures S4 and S5). This supports our choice of adopting the small-mutation approximation described in the previous paragraph.

References

1. Van Lange PAM, Balliet D, Parks CD, Van Vugt M. Social Dilemmas – The Psychology of Human Cooperation. Oxford: Oxford University Press; 2015.
2. Kollock P. Social dilemmas: The anatomy of cooperation. *Annu Rev Sociol.* 1998;24:183–214.
3. Dawes RM. Social dilemmas. *Annu Rev Psychol.* 1980;31(1):169–93.
4. Rand DG, Nowak MA. Human cooperation. *Trends Cogn Sci.* 2013 Aug;17(8):413–25.
5. Trivers RL. The evolution of reciprocal altruism. *Q Rev Biol.* 1971;46:35–57.
6. Axelrod R. *The Evolution of Cooperation*. New York, NY: Basic Books; 1984.
7. Nowak MA. Five rules for the evolution of cooperation. *Science.* 2006;314(5805):1560–3.
8. Schmid L, Chatterjee K, Hilbe C, Nowak MA. A unified framework of direct and indirect reciprocity. *Nat Hum Behav.* 2021 Oct 1;5(10):1292–302.
9. Doebeli M, Hauert C. Models of cooperation based on the Prisoner’s Dilemma and the Snowdrift game. *Ecol Lett.* 2005 Jul;8(7):748–66.
10. Stewart AJ, Plotkin JB. From extortion to generosity, evolution in the Iterated Prisoner’s Dilemma. *Proc Natl Acad Sci U S A.* 2013 Oct 17;110(38):15348–53.
11. Imhof LA, Fudenberg D, Nowak MA. Tit-for-tat or win-stay, lose-shift? *J Theor Biol.* 2007 Aug 7;247(3):574–80.

12. Nowak MA, Sigmund K. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*. 1993 Jun 1;364:56–8.
13. Van Veelen M, García J, Rand DG, Nowak MA. Direct reciprocity in structured populations. *Proc Natl Acad Sci U S A*. 2012 Jun 19;109(25):9929–34.
14. Vallet A, Locatelli B, Barnaud C, Makowski D, Quispe Conde Y, Levrel H. Power asymmetries in social networks of ecosystem services governance. *Environ Sci Policy*. 2020 Dec 1;114:329–40.
15. Scheffer M, Van Bavel B, Van De Leemput IA, Van Nes EH. Inequality in nature and society. *Proc Natl Acad Sci U S A*. 2017 Dec 12;114(50):13154–7.
16. Piketty T, Saez E. Inequality in the long run. *Science*. 2014;344(6186):838–43.
17. Gerpott FH, Balliet D, Columbus S, Molho C, de Vries RE. How do people think about interdependence? A multidimensional model of subjective outcome interdependence. *J Pers Soc Psychol*. 2018 Oct 1;115(4):716–42.
18. Redhead D, Power EA. Social hierarchies and social networks in humans. *Phil Trans R Soc B*. 2022 Feb 28;377(1845):20200440.
19. Hall JA, Coats EJ, LeBeau LS. Nonverbal behavior and the vertical dimension of social relations: A meta-analysis. *Psychol Bull*. 2005 Nov;131(6):898–924.
20. Fiske ST, Berdahl JL. Social power. In: Kruglanski AW, Tory Higgins E, editors. *Social psychology: Handbook of basic principles*. New York: The Guilford Press; 2007. p. 678–92.
21. Smith PK, Hofmann W. Power in everyday life. *Proc Natl Acad Sci U S A*. 2016 Sep 6;113(36):10043–8.
22. Kaufmann JH. On the definitions and functions of dominance and territoriality. *Biological Reviews*. 1983;58(1):1–20.
23. De Vries H, Stevens JMG, Vervaecke H. Measuring and testing the steepness of dominance hierarchies. *Anim Behav*. 2006 Mar;71(3):585–92.
24. Tibbetts EA, Pardo-Sanchez J, Weise C. The establishment and maintenance of dominance hierarchies. *Phil Trans R Soc B*. 2022 Feb 28;377(1845).
25. Hammerstein PB. The role of asymmetry in animal contests. *Anim Behav*. 1981;29:193–205.
26. Gaunersdorfer A, Hofbauer J, Sigmund K. On the Dynamics of Asymmetric Games. *Theor Popul Biol*. 1991;39(3):345–57.
27. Maynard-Smith J, Parker GA. The logic of asymmetric contest. *Anim Behav*. 1976;24:159–75.
28. Dawkins MS. Do asymmetries destabilize the Prisoner's Dilemma and make reciprocal altruism unlikely? *Anim Behav*. 2010 Aug;80(2):339–41.
29. Hauser OP, Hilbe C, Chatterjee K, Nowak MA. Social dilemmas among unequals. *Nature*. 2019 Aug 22;572(7770):524–7.

30. Ladret V, Lessard S. Evolutionary game dynamics in a finite asymmetric two-deme population and emergence of cooperation. *J Theor Biol.* 2008 Nov 7;255(1):137–51.
31. Thibaut JohnW, Kelley HH. *The social psychology of groups.* New York: John Wiley; 1959.
32. Emerson RM. Power-dependence relations. *Am Sociol Rev.* 1962;27:31–41.
33. Ahn TK, Lee M, Ruttan L, Walker J. Asymmetric payoffs in simultaneous and sequential prisoner's dilemma games. *Public Choice.* 2007 Sep;132(3–4):353–66.
34. Beckenkamp M, Frank HHS, Maier-Rigaud P, Hennig-Schmidt H, Maier-Rigaud FP, Engel C, et al. Cooperation in symmetric and asymmetric prisoner's dilemma games. *MPI Collective Goods Preprint.* 2006;25.
35. Sheposh JP, Gallo PS. Asymmetry of payoff structure and cooperative behavior in the prisoner's dilemma game. *Journal of Conflict Resolution.* 1973;17(2):321–33.
36. Talley MB. *Effects of asymmetry of payoff and asymmetry of information in a prisoner's dilemma game.* Arlington: University of Texas; 1974.
37. Bone JE, Wallace B, Bshary R, Raihani NJ. Power asymmetries and punishment in a prisoner's dilemma with variable cooperative investment. *PLoS One.* 2016 May 1;11(5).
38. Nikiforakis N, Normann HT, Wallace B. Asymmetric enforcement of cooperation in a social dilemma. *South Econ J.* 2010 Jan;76(3):638–59.
39. Hilbe C, Hagel K, Milinski M. Asymmetric power boosts extortion in an economic experiment. *PLoS One.* 2016 Oct 1;11(10).
40. Molho C, Balliet D, Wu J. Hierarchy, power, and strategies to promote cooperation in social dilemmas. *Games (Basel).* 2019 Mar 1;10(1).
41. Kopelman S. The effect of culture and power on cooperation in commons dilemmas: Implications for global resource management. *Organ Behav Hum Decis Process.* 2009 Jan;108(1):153–63.
42. Halevy N, Y. Chou E, D. Galinsky A. A functional model of hierarchy. *Organizational Psychology Review.* 2011 Feb;1(1):32–52.
43. Antonioni A, Pereda M, Cronin KA, Tomassini M, Sánchez A. Collaborative hierarchy maintains cooperation in asymmetric games. *Sci Rep.* 2018 Dec 1;8(1).
44. French JR, Raven BH. The bases of social power. In: Cartwright D, editor. *Studies in social power.* Ann Arbor: Univeristy of Michigan; 1959. p. 150–67.
45. Keltner D, Gruenfeld DH, Anderson C. Power, approach, and inhibition. *Psychol Rev.* 2003;110(2):265.
46. Balliet D, Tybur JM, Van Lange PAM. Functional interdependence theory: An evolutionary account of social situations. *Personality and Social Psychology Review.* 2017 Nov 1;21(4):361–88.
47. Kelley HH, Holmes JG, Kerr NL, Reis HT, Rusbult CE, Van Lange PAM. *An Atlas of Interpersonal Situations.* Cambridge: Cambridge University Press; 2003.

48. Carney DR. The nonverbal expression of power, status, and dominance. *Curr Opin Psychol.* 2020 Jun 1;33:256–64.
49. Aguinis H, Simonsen MM, Pierce CA. Effects of nonverbal behavior on perceptions of power bases. *Journal of Social Psychology.* 1998 Aug 1;138(4):455–69.
50. Carney DR, Hall JA, LeBeau LS. Beliefs about the nonverbal expression of social power. *J Nonverbal Behav.* 2005 Jun;29(2):105–23.
51. Brey E, Shutts K. Children use nonverbal cues to make inferences about social power. *Child Dev.* 2015 Jan 1;86(1):276–86.
52. Langner CA, Keltner D. Social power and emotional experience: Actor and partner effects within dyadic interactions. *J Exp Soc Psychol.* 2008 May;44(3):848–56.
53. Berdahl JL, Martorana P. Effects of power on emotion and expression during a controversial group discussion. *Eur J Soc Psychol.* 2006 Jul;36(4):497–509.
54. Colnaghi M, Santos FP, Van Lange PAM, Balliet D. Adaptations to infer fitness interdependence promote the evolution of cooperation. *Proc Natl Acad Sci U S A.* 2023 Dec 12;120(50).
55. Maynard-Smith J. The evolution of behavior. *Sci Am.* 1978;239(3):176–93.
56. Skyrms B. *The stag hunt and the evolution of social structure.* Cambridge University Press; 2004.
57. Halevy N, Chou EY, Murnighan JK. Mind games: The mental representation of conflict. *J Pers Soc Psychol.* 2012 Jan;102(1):132–48.
58. Santos FC, Pacheco JM, Lenaerts T. Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc Natl Acad Sci U S A.* 2006;103:3490–4.
59. Peña J, Nöldeke G. Cooperative dilemmas with binary actions and multiple players. *Dyn Games Appl.* 2023;13:1156–93.
60. Bramoullé Y. Anti-coordination and social interactions. *Games Econ Behav.* 2007 Jan 1;58(1):30–49.
61. King AJ, Johnson DDP, Van Vugt M. The origins and evolution of leadership. *Current Biology.* 2009 Oct 13;19(19).
62. Van Vugt M, Hogan R, Kaiser RB. Leadership, followership, and evolution: some lessons from the past. *American Psychologist.* 2008 Apr;63(3):182–96.
63. Pietraszewski D. The evolution of leadership: Leadership and followership as a solution to the problem of creating and executing successful coordination and cooperation enterprises. *Leadersh Q.* 2020 Apr 1;31(2):101299.
64. Sigmund K. *The Calculus of Selfishness.* Princeton University Press; 2010.
65. Baek SK, Jeong HC, Hilbe C, Nowak MA. Comparing reactive and memory-one strategies of direct reciprocity. *Sci Rep.* 2016 May 10;6.

66. Maynard-Smith J. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press; 1982.
67. Fudenberg D, Imhof LA. Imitation processes with small mutations. *J Econ Theory*. 2006 Nov 1;131(1):251–62.
68. Eckel CC, Wilson RK. Social learning in coordination games: Does status matter? *Exp Econ*. 2007 Sep;10(3):317–29.
69. Ball SB, Eckel CC. Buying status: Experimental evidence on status in negotiation. *Psychol Mark*. 1996;13(4):381–405.
70. Kwaadsteniet EW de, Dijk E van. Social status as a cue for tacit coordination. *J Exp Soc Psychol*. 2010 May;46(3):515–24.
71. Gonzalez C, Ben-Asher N, Martin JM, Dutt V. A cognitive model of dynamic cooperation with varied interdependency information. *Cogn Sci*. 2015 Apr 1;39(3):457–95.
72. Imhof LA, Fudenberg D, Nowak MA. Evolutionary cycles of cooperation and defection. *Proc Natl Acad Sci U S A*. 2005;102(31):10797–800.
73. Taha MA, Ghoneim A. Zero-determinant strategies in repeated asymmetric games. *Appl Math Comput*. 2020 Mar 15;369.
74. Press WH, Dyson FJ. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proc Natl Acad Sci U S A*. 2012 Jun 26;109(26):10409–13.
75. Godfray HCJ. The evolution of forgiveness. *Nature*. 1992 Jan 16;355:206–7.
76. Couto MC, Giaimo S, Hilbe C. Introspection dynamics: A simple model of counterfactual learning in asymmetric games. *New J Phys*. 2022 Jun 1;24(6).
77. Pauwels L, Declerck CH, Boone C, Diaz-Gutiérrez P, Lambert B. Does Power Corrupt? An fMRI Study on the Effect of Power and Social Value Orientation on Inequity Aversion. *J Neurosci Psychol Econ*. 2022;15(4):222–40.
78. Moran PAP. *The statistical process of evolutionary theory*. Oxford: Clarendon Press; 1962.
79. Traulsen A, Shorish N, Nowak MA. Analytical results for individual and group selection of any intensity. *Bull Math Biol*. 2008 Jul;70(5):1410–24.
80. Nowak MA, Sasaki A, Taylor C, Fudenberg D. Emergence of cooperation and evolutionary stability in finite populations. *Nature*. 2004 Apr 8;428(6983):646–50.

FIGURES

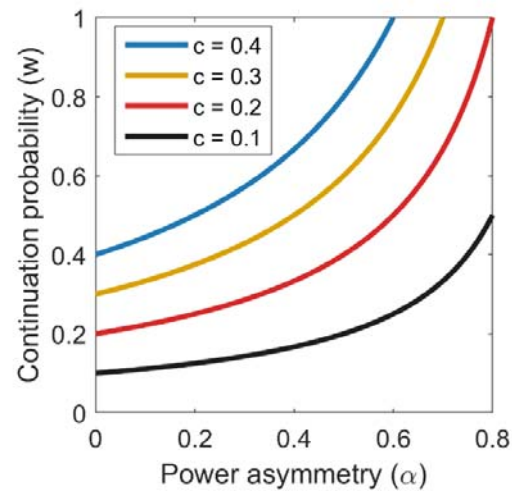


Figure 1. Threshold continuation probability. Minimum continuation probability w^* that makes reciprocal cooperation (TFT) evolutionarily stable. Increasing power asymmetry (α) or the cost of cooperation (c) destabilizes cooperation. Other parameters: $b = 1$.

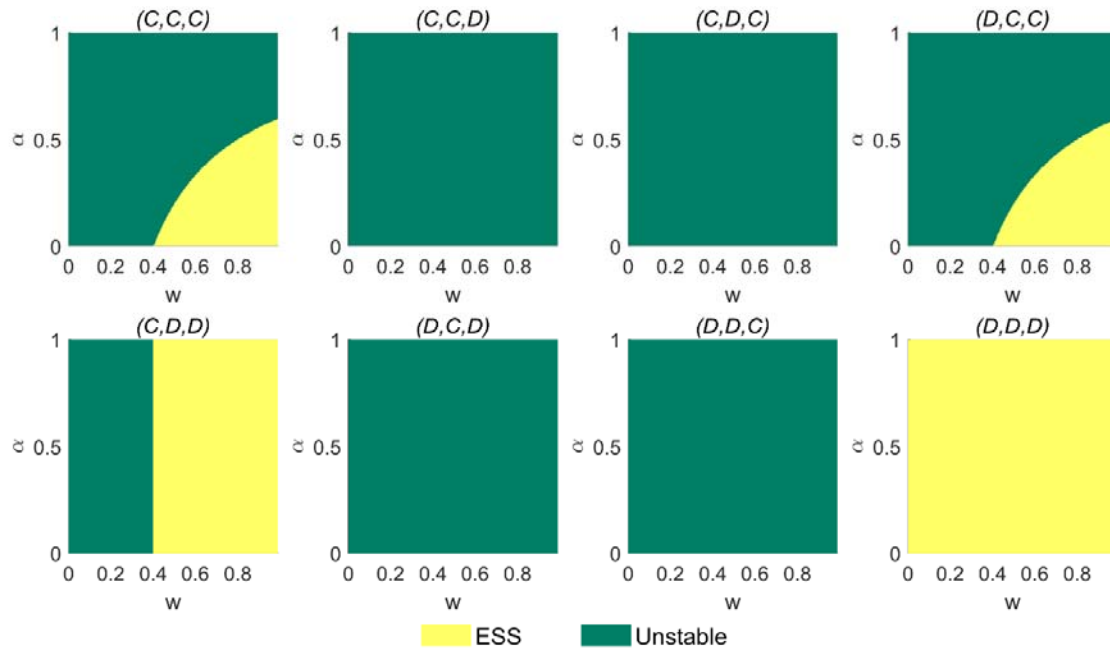


Figure 2. Evolutionarily stable strategies in the iterated PD. Evolutionarily stable strategies in the iterated Prisoner's Dilemma for varying levels of power asymmetry (α) and continuation probability (w). The area of the parameter space where each strategy is an ESS is indicated in yellow. Other parameters: $b = 1$, $c = 0.4$.

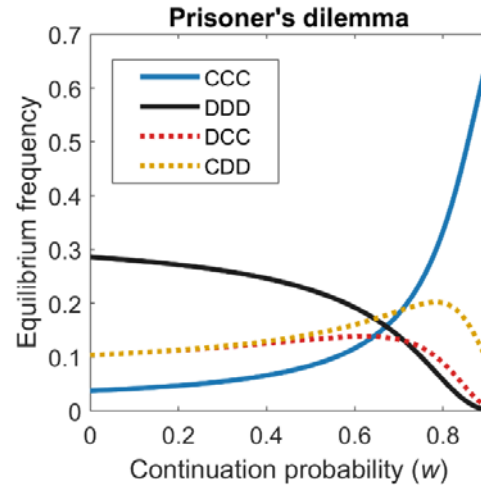


Figure 3. Equilibrium frequency of strategies in the iterated PD. Equilibrium frequency of various strategies in a finite population playing the iterated PD, under a small-mutation approximation, as a function of the continuation probability (w). Continuous lines indicate power-independent strategies; dotted lines indicate strategies that are conditional on power differences. Of the 8 strategies studied, only the four ESSs are shown. Other parameters: $\alpha = 0.5$, $b = 1$, $c = 0.4$, $N = 100$, $\beta = 0.05$.

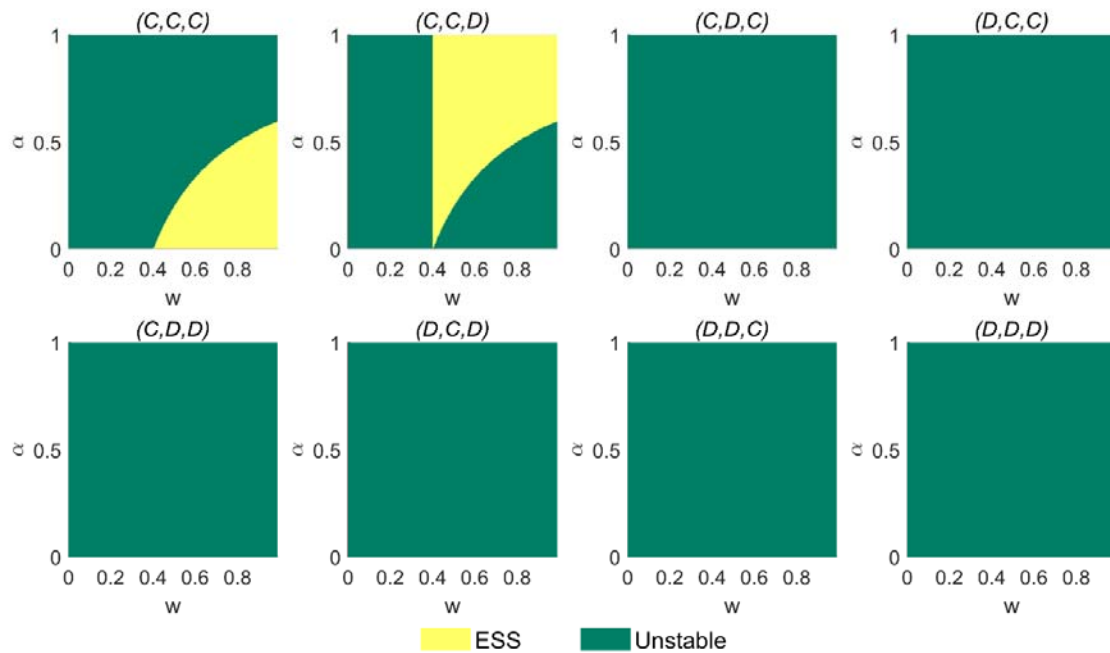


Figure 4. Evolutionarily stable strategies in the iterated SD. Evolutionarily stable strategies in the iterated Snowdrift for varying levels of power asymmetry (α) and continuation probability (w). The area of the parameter space where each strategy is an ESS is indicated in yellow. Other parameters: $b = 1$, $c = 0.4$.

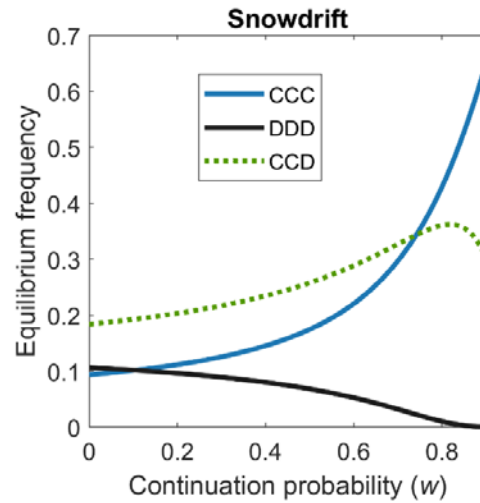


Figure 5. Equilibrium frequency of strategies in the iterated SD. Equilibrium frequency of various strategies in a finite population playing the iterated SD, under a small-mutation approximation, as a function of the continuation probability (w). Continuous lines indicate power-independent strategies; the dotted line indicates (C, D, D) , a strategy that is conditional on power differences. Of the 8 strategies studied, only the two ESSs and AIID are shown. Other parameters: $\alpha = 0.5$, $b = 1$, $c = 0.4$, $N = 100$, $\beta = 0.05$.

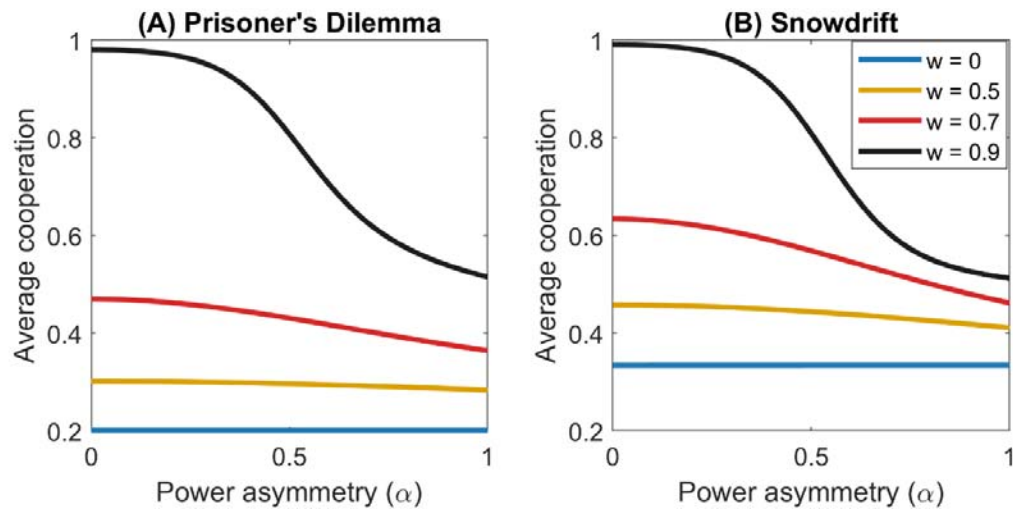


Figure 6. Average frequency of cooperation. Average frequency of cooperation in a finite population under a small-mutation approximation, as a function of power asymmetry (α) and the continuation probability (w), in the iterated PD (A) and SD (B). Other parameters: $b = 1$, $c = 0.4$, $N = 50$, $\beta = 0.1$.