



Successful strategies in the voluntarily repeated Prisoner's Dilemma

Luis R. Izquierdo ^{a,*}, Segismundo S. Izquierdo ^b, Robert Boyd ^c

^a Department of Management Engineering, University of Burgos, Burgos, 09001, Spain

^b Department of Industrial Organization and BioEcoUva, University of Valladolid, Valladolid, 47011, Spain

^c School of Human Evolution and Social Change, Arizona State University, Tempe, AZ, 85287, USA

ARTICLE INFO

Keywords:

Cooperation
Evolutionary game theory
Prisoner's Dilemma
Voluntary separation
Option to leave
Positive assortment

ABSTRACT

Reciprocity is central to explanations of cooperation among unrelated individuals in societies of humans and other animals. Most mathematical analyses of the evolution of reciprocity are based on the repeated Prisoner's Dilemma (RPD) and typically assume that new strategies are rarely introduced by mutation or analogous cultural processes, that behavioral errors are absent or infrequent, and that agents are bound to interact with the same partner. Here we analyze a version of the RPD in which new strategies are frequently tested, behavioral errors occur at a non-negligible rate, and actors may have the option to leave their current partner. In this environment, the usual indeterminacy disappears, and the mix of strategies and cooperation levels are quite stable. With the option to leave, cooperation persists at a substantially higher level than without the option to leave. Classical strategies such as *Tit-for-Tat*, *Win-Stay-Lose-Shift*, and *Grim* become rare and are largely displaced by strategies that sanction cheaters by leaving rather than by retaliatory defection. Beyond a threshold, increasing the number of times partners interact decreases the level of cooperation without the option to leave, but increases it when leaving is possible.

1. Introduction

Cooperation among unrelated individuals is a defining feature of human societies and a central puzzle across the biological and social sciences. The repeated Prisoner's Dilemma (RPD) has long served as the workhorse model to explain how pairwise reciprocity can emerge. By conditioning future actions on past behavior, individuals can sustain mutually beneficial cooperation.

The theoretical study of the RPD within classical game theory has largely been guided by an equilibrium-based perspective, in which predictions are framed in terms of equilibrium concepts such as Nash equilibrium and their refinements. A central difficulty in this approach is indeterminacy: for sufficiently large continuation probability, the Folk Theorem implies the existence of infinitely many Nash equilibria in the RPD. However, none of them are strict. From an evolutionary perspective, there are infinitely many neutrally stable strategies (Bendor and Swistak, 1995, 1998), but none are evolutionarily stable (Boyd and Loberbaum, 1987) or dynamically stable (García and van Veelen, 2016). Thus, it seems impossible to identify "successful" strategies since, depending on the definition of success, there are infinitely many or there are none.

Introducing behavioral errors—so that players may occasionally choose unintended actions—alters this picture but fails to resolve the

indeterminacy. With rare errors, a large number of strict Nash strategies appear (Sugden, 1986; Boyd, 1989; Kim, 1994; Leimar, 1997), many of which can support full cooperation. These equilibria correspond to asymptotically stable states under most evolutionary dynamics, so the fundamental problem of identifying a clear "winning" behavior remains.

To obtain sharper predictions, a common approach has been to study evolutionary models with vanishingly rare mutations (Kandori et al., 1993; Young, 1993, 1998; Fudenberg and Imhof, 2006; Fudenberg et al., 2006; Traulsen and Hauert, 2009; Wu et al., 2012; Hilbe et al., 2013; Stewart and Plotkin, 2013; Stewart et al., 2016; García and van Veelen, 2016; van Veelen and García, 2019; Schmid et al., 2021). In these models, populations spend most of their time in homogeneous states, with evolutionary change occurring through rare transitions between these monomorphic states, as mutants either go extinct or fixate before new mutations can arise. Long-run behavior is then characterized by the relative prevalence of equilibria, reinforcing the equilibrium-selection perspective that dominates the theoretical analysis of the RPD.

A central limitation of this body of work is the focus on settings with negligible or nonexistent errors and mutation—an assumption that clashes with the noisy reality of biological and social systems. Empirical evidence from both biological and social systems indicates that behavioral errors and mutation are common, and that substantial variation

* Corresponding author.

E-mail addresses: lizquierdo@ubu.es (L.R. Izquierdo), segismundo.izquierdo@uva.es (S.S. Izquierdo), rboyd3@asu.edu (R. Boyd).

<https://doi.org/10.1016/j.jtbi.2026.112598>

Received 6 June 2026; Received in revised form 13 August 2026; Accepted 17 September 2026

Available online 19 September 2026

0022-5193/© 2026 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

is persistently maintained within populations (Grujić et al., 2014; Boyd and Mathew, 2021; Tkadlec et al., 2023). While single-locus mutation rates in standard genetic models are typically low, complex behavioral strategies are influenced by many loci with small effects, which allows substantial heritable genetic variation to exist at equilibrium (Chabris et al., 2015). In this context, high mutation rates serve as a practical and effective proxy for modeling this underlying standing variation (e.g., Kokko et al., 2002). Furthermore, in human social contexts, behavioral variation arises naturally through continuous individual experimentation and errors in strategy imitation. Evidence from social learning experiments and naturalistic observations confirms that such innovation and error rates are indeed substantial (Traulsen et al., 2010; Grujić et al., 2014; Tkadlec et al., 2023).

The importance of variation is not merely empirical: it can alter the qualitative predictions of evolutionary games. In a finitely repeated Prisoner's Dilemma, extrinsically maintained behavioral variation can stabilize high levels of cooperation even when the corresponding homogeneous model predicts defection (McNamara et al., 2004). More broadly, even small amounts of maintained variation can change strategic incentives by preserving the value of learning about others and responding contingently to their behavior (McNamara and Leimar, 2010). In evolutionary models, recurrent mutation is one mechanism that maintains such variation. When mutation is non-negligible, the resulting dynamics need not be a small perturbation of the rare-mutation limit but may instead constitute a qualitatively distinct dynamical regime (Imhof et al., 2005; Hilbe et al., 2017; Tkadlec et al., 2023). A striking consequence of allowing substantial mutation and errors is that predictions often become sharper. When mutation and error rates are large enough to maintain persistent variation, the profusion of equilibria characteristic of repeated games collapses to a small number of dynamical attractors (Imhof et al., 2005). This collapse makes it possible to move beyond equilibrium enumeration and instead identify the behavioral traits that characterize successful strategies, as well as the levels of cooperation sustained in the long run.

In addition to the standard RPD, we consider the voluntarily repeated Prisoner's Dilemma (VRPD), in which players may terminate partnerships and seek a new partner (Friedman and Hammerstein, 1991; Enquist and Leimar, 1993; Noë and Hammerstein, 1995; Hayashi and Yamagishi, 1998; Aktipis, 2004; Hamilton and Taborsky, 2005; Fujiwara-Greve and Okuno-Fujiwara, 2009; Izquierdo et al., 2010; Aktipis, 2011; Barclay, 2013; Barclay and Raihani, 2016; Graser et al., 2025; Gutiérrez-Mielgo et al., 2025; Izquierdo and Izquierdo, 2026). This setting captures situations in which individuals can leave uncooperative partners, such as friendships, marriages, economic relationships, and mutual aid in humans and other animals (Mathew and Boyd, 2026). Equilibrium analysis of the VRPD is even less selective than in the RPD: many partially cooperative equilibria exist (Izquierdo and Izquierdo, 2026), and evolutionary models with rare mutation likewise fail to identify a small set of successful strategies (Graser et al., 2025). As in the RPD, equilibrium multiplicity persists even in the presence of partner choice, and neither equilibrium analysis nor evolutionary models with rare mutation are sufficient to explain which behaviors may be observed. Few models have considered the option to leave with substantial variation. A notable exception is McNamara et al. (2008), who elegantly demonstrated that voluntary separation can drive the evolution of cooperation when populations are subject to substantial variation. However, in that framework, individuals are characterized by genetically fixed levels of cooperativeness; they can respond to a partner's actions by leaving, but they cannot adjust their cooperativeness within an ongoing partnership.

A related line of work represents opting out through strategy-dependent interaction durations. These models show that sufficiently persistent interactions between cooperators can support a stable state with a high frequency of cooperation, although all-defection remains locally stable (Křivan and Cressman, 2017). This approach has also been investigated experimentally and developed in graph-structured and multiplayer settings (Zhang et al., 2016; Kurokawa, 2019; Broom

and Křivan, 2020; Křivan and Cressman, 2020). These studies provide complementary examples of how selective interaction duration can generate assortment, but their main focus is on how particular duration or separation patterns affect cooperation within restricted strategy sets. By contrast, in our model, partnership duration is endogenously shaped by evolving conditional separation decisions rather than assigned to particular pair types in advance.

More generally, previous models lack substantial variation, omit execution noise, or prohibit leaving a partner by design. Consequently, they cannot evaluate how classical reciprocal behaviors—which depend on precise coordination of conditional responses—fare under realistic, noisy conditions where agents can leave their partners.

Here we address exactly this gap by comparing the RPD and the VRPD in an evolutionary framework with *both* substantial behavioral errors and mutation. We focus on deterministic memory-one strategies (Hilbe et al., 2017), which include most of the widely studied strategies in the literature while avoiding unbounded strategic complexity. In this setting, evolutionary trajectories in both games converge to a small number of robust regimes characterized by stable cooperation levels and well-defined behavioral traits. This allows us to identify which strategies are successful and to compare cooperation with and without the option to leave.

We find that, despite the absence of fully cooperative equilibria in the VRPD, cooperation is higher and more robust in the VRPD than in the RPD due to positive assortment generated by partner switching. This result aligns with experimental evidence (Fehl et al., 2011; Rand et al., 2011; Barclay and Raihani, 2016), which shows that cooperation is higher when individuals can leave partners and that partner switching is frequent when costs are not excessive. Our results also show that classical strategies—such as Tit-for-Tat (*TFT*) and Win-Stay-Lose-Shift (*WSLS*)—that have been celebrated for their success in the RPD (Axelrod, 1984; Kraines and Kraines, 1993; Nowak and Sigmund, 1993; Imhof et al., 2005) perform poorly in the VRPD: with moderate error and mutation rates, these strategies are practically eliminated from the population. More broadly, our findings reveal that when social interactions are subject to realistic behavioral noise, the freedom to walk away is a far more robust driver of cooperation than the threat of retaliation.

2. The model

Our model encompasses the standard RPD and the VRPD. The standard RPD is obtained as a special case of the VRPD by forbidding strategies that use the option to leave. Fig. 1 shows a sketch of the sequence of events that take place in the model within each (discrete) time step, which we call *tick*. At the beginning of every tick, all single players (those without a partner) are randomly matched in partnerships, and then all agents play the stage game. Analysis presented in the SI indicates that our results are not substantially affected by delayed rematching.

The stage game is a Prisoner's Dilemma (PD). Actors can either cooperate (*C*) or defect (*D*) with payoffs $U_{CC} = b$, $U_{CD} = 0$, $U_{DC} = b + 1$ and $U_{DD} = 1$, where the first action is the actor's and the second their partner's. This version of the PD is often called the *donation game*, where $b > 1$ is the benefit provided by a cooperating partner at a cost equal to 1.

Each player's strategy determines their action given the partnership's history. We consider memory-one pure strategies, i.e., deterministic strategies whose choices depend only on the partnership's last outcome. We can represent any of these strategies by a sequence of five letters indicating what choice to take (*C*, *D* or *L*, where *L* denotes leaving the partnership) after each possible history ($\emptyset, CC, CD, DC, DD$). For example, $C|CLCC$ is the strategy that starts a new partnership playing *C*, never plays *D*, and leaves after a partner's unilateral defection. There are $2 \times 3^4 = 162$ memory-one strategies in the VRPD. The memory-one strategies for the RPD are the $2 \times 2^4 = 32$ strategies with no *L*. For instance, *Tit-for-Tat* (*TFT*) corresponds to $C|CDCD$, while $C|CDDD$ is a strategy that chooses to cooperate with new partners and keeps on

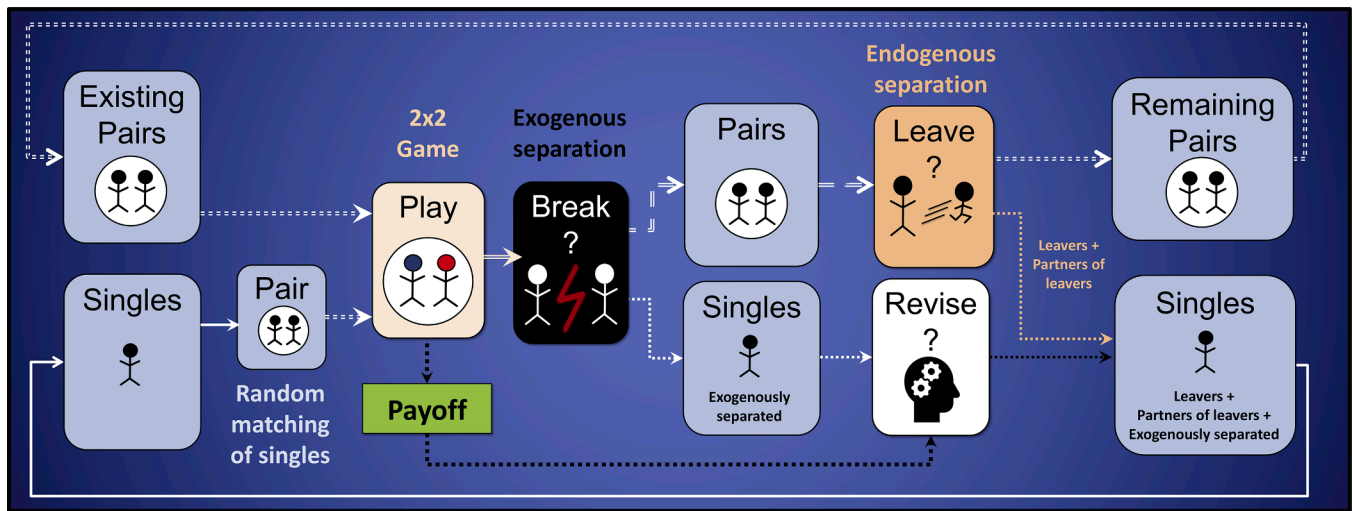


Fig. 1. Sketch of the sequence of events within each discrete time step, or *tick*. The “Remaining Pairs” and “Singles” at the end of a time step are identical to the “Existing Pairs” and “Singles” in the next time step.

cooperating if and only if there is mutual cooperation, defecting otherwise. In a slight abuse of convention, we call this last strategy *Grim*.

Errors are modeled as unintended actions in the stage game: a *D* instead of an intended *C*, or a *C* instead of an intended *D*. Both errors occur with probability ϵ . After playing the stage game, partnerships may dissolve for exogenous reasons. The probability that this does not occur and the partnership continues is δ . Thus, $\gamma = (1 - \delta)^{-1}$ is the *expected number of interactions* of partnerships whose members do not choose to break up.

After every exogenous breakup, players may revise their strategy independently, with probability r . Revisers experiment with probability μ —which we refer to as the *mutation* probability to maintain consistency with the evolutionary game theory literature—and imitate with probability $1 - \mu$. Under our local experimentation rule, one of the five components of the current strategy is selected uniformly at random and its prescribed action is replaced by one of the other feasible actions, also selected uniformly. Thus, experimentation alters behavior following a single contingency rather than replacing the entire strategy. Further details and discussion of this mutation mechanism are provided in the SI.

When revisers imitate, they adopt a strategy that is currently present in the population. Each strategy is selected with probability proportional to the total payoff earned during the current tick by all individuals using that strategy. This procedure is exactly equivalent to selecting an individual from the population with probability proportional to that individual’s current payoff and copying their strategy. This updating process can be interpreted as cultural imitation, migration, or reproduction with mutation.

Simulations were run in finite populations of constant size N . Each simulation was run for sufficiently many ticks for the reported summary statistics to stabilize after the initial transient; the exact simulation horizons and averaging windows are stated in the relevant figure and table captions. We also used a deterministic mean-dynamic approximation to study the corresponding dynamics in large populations. Further implementation details and replication resources are provided in the SI.

3. Simulation results

In this section, we use computer simulations to compare the RPD and the VRPD. At the beginning of each simulation run, all individuals are single. Each individual’s initial strategy is drawn independently and uniformly from the relevant strategy set: the 32 strategies in the RPD or the 162 strategies in the VRPD. Thus, every feasible strategy has the

same probability of being assigned initially, although the realized initial frequencies vary across runs.

Our baseline parameterization is $b = 3$, $\delta = 0.95$ ($\gamma = 20$), $\epsilon = 0.05$, $\mu = 0.05$, $r = 0.1$, and $N = 500$. These values are intended as a theoretically informative benchmark rather than an empirically calibrated scenario. In particular, $\epsilon = 0.05$ represents a moderate level of behavioral noise, corresponding on average to one unintended action in every 20 choices, while $\gamma = 20$ represents a relationship expected to last 20 interactions in the absence of endogenous separation. These values ensure that both behavioral errors and the consequences of repeated interaction are relevant. As shown below, our main qualitative findings are robust to varying each parameter individually over a broad range of values.

In the baseline scenario, allowing individuals to leave their partners raises the estimated long-run cooperation rate from less than 40% to more than 60%. More precisely, the long-run mean proportion of *CC* outcomes is 24.5 percentage points higher in the VRPD than in the RPD (95% Welch confidence interval for the difference: [22.5, 26.5] percentage points).

Fig. 2 shows how the long-run cooperation rate changes when each parameter is varied individually while all other parameters are held at their baseline values. This one-at-a-time analysis evaluates robustness to each parameter separately. Specifically, each point in Fig. 2 reports the mean cooperation rate from tick 190 000 to tick 200 000 over 100 runs starting from random initial conditions (i.e., initially, each agent adopts a random strategy). The error bars show pointwise 95% *t*-confidence intervals for the corresponding mean. In many cases, these intervals are narrower than the plotting symbols. Overall, the cooperation advantage of the VRPD is robust across a broad range of parameter values, although it does not hold at every value examined (e.g., payoff benefit $b < 2$, expected number of interactions $\gamma \leq 7$ or $\epsilon = 0$). Accordingly, within the parameter range examined, the cooperation advantage of the VRPD arises when the benefit of cooperation is sufficiently large and relationships are noisy and persist long enough for contingent behavior to matter; it is absent when the benefit is small, interactions are error-free, or the expected number of encounters is very small.

Fig. 2(b) shows how cooperation is affected by the expected number of interactions, γ . If $\gamma \leq 3$, cooperation is not observed: when partnerships last such a short time, the short-lived rewards enjoyed by partnerships of cooperators cannot compensate for exploitation by defectors. Cooperation first becomes appreciable at $\gamma = 4$. For $\gamma \in \{4, 5, 6, 7\}$, mean cooperation is higher in the RPD than in the VRPD. Then, for larger values of γ , an interesting pattern appears: increasing the

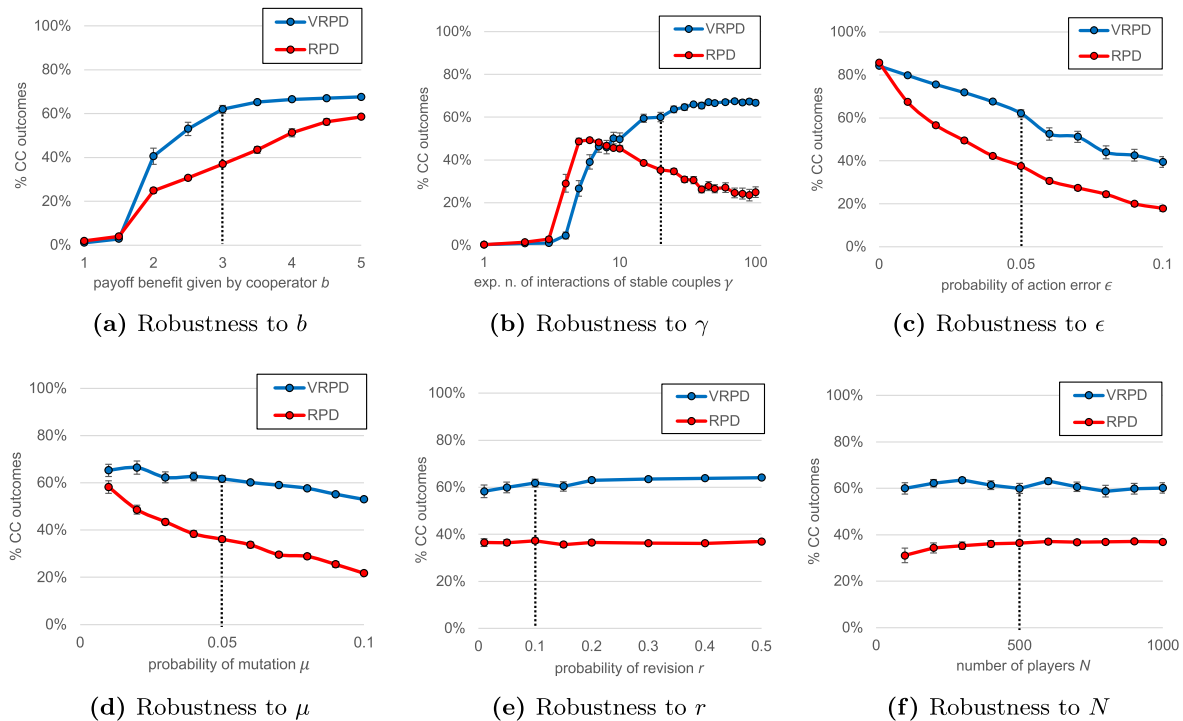


Fig. 2. Average percentage of cooperative outcomes in the RPD and VRPD when each parameter is varied from its baseline value: (a) the payoff benefit b , (b) the expected number of interactions of partnerships that are not voluntarily terminated γ , (c) the probability of action error ϵ , (d) the probability of mutation μ , (e) the probability of revision r , and (f) the population size N . Each point is the average from tick 190 000 to tick 200 000 over 100 runs starting from random initial conditions. The dashed vertical line in each plot indicates the value of the parameter in the baseline scenario. The error bars show pointwise 95% t -confidence intervals; the widest confidence interval has a width below 9 percentage points. In panel (b), the mean curves cross between $\gamma = 7$ and $\gamma = 8$, marking a transition from cooperation being higher when leaving is unavailable (RPD) to cooperation being higher when leaving is allowed (VRPD). In the RPD, cooperation reaches its maximum at approximately $\gamma = 6$ and subsequently declines, whereas cooperation continues to increase over a wider range of values when leaving is allowed. Without the option to leave (RPD), longer relationships allow errors to generate prolonged retaliation or miscoordination; with the option to leave (VRPD), individuals can abandon such partnerships and be rematched. The crossing and turning points are descriptive features of the plotted means rather than statistically estimated thresholds.

Table 1

Percentage of players using each choice (row) after a given history (column) in the RPD, in the baseline scenario. Values are averages from tick 290 000 to tick 300 000 over 100 runs starting from random initial conditions.

		Empty history $\emptyset$	Previous outcome			
			CC	CD	DC	DD
Choice	C	87	91	9	18	38
	D	13	9	91	82	62

number of expected interactions (γ) lowers cooperation in the RPD, but increases cooperation in the VRPD. Thus, the cooperation gap between the RPD and VRPD widens as the number of interactions increases.

To understand this, we examine which strategies prevail in the baseline scenario for the RPD and for the VRPD. Table 1 shows that in the RPD selection favors: (1) being nice (87%), (2) cooperating after mutual cooperation (91%), (3) defecting after exploitation (91%), and (4) continuing to defect after exploiting (82%). The only two (memory-one) strategies with these four traits are *Grim* and *WSLS*. The difference between these two strategies is that *Grim* defects after mutual defection, while *WSLS* cooperates.

Fig. 3 shows the distribution across runs of the mean frequencies of the four most prevalent strategies in the RPD simulations in the baseline scenario, together with their aggregate distribution (rightmost box plot). They are all nice, continue cooperating after mutual cooperation,

and defect after being exploited, i.e., they exhibit the pattern $C|CD\circ\circ$ (where $\circ$ may be C or D). At baseline, this set of strategies accounts for an average of 72.2% of the RPD population (95% t -confidence interval: [70.8, 73.6]%). In the absence of errors, any match between two members of this set leads to full cooperation.

Importantly, although the aggregate frequency of the set $C|CD\circ\circ$ is high and relatively stable across runs (see the rightmost box plot in Fig. 3), there are strong fluctuations over time and across runs in the frequency of each individual strategy (see Fig. 4). We therefore do not interpret the apparent differences among the individual box plots as systematic differences in evolutionary success. Instead, our analysis highlights the collective prevalence and shared behavior of the strategy set $C|CD\circ\circ$.

The presence of errors is crucial for understanding why increasing the expected number of interactions γ can lead to less cooperation (Fig. 2(b)). In the RPD, the only way cooperative strategies can protect themselves from exploiters is by defecting, and selection favors this trait: most players defect after CD (see Table 1). With non-negligible errors, however, this means that the longer a partnership lasts, the more likely it is that there will be mistakes that lead to sequences of mutual defection or miscoordination. Moreover, increasing γ reduces the frequency of exogenous breakup and hence the opportunities for such sequences to be interrupted by rematching and a new partnership history. Thus, in the RPD, intermediate partnership lengths lead to the highest cooperation rates: when errors are non-negligible, the expected number of interactions should be neither too low nor too high.

Now consider the VRPD. The possibility of leaving a partner introduces two important features. First, players can use the *defect-and-leave*

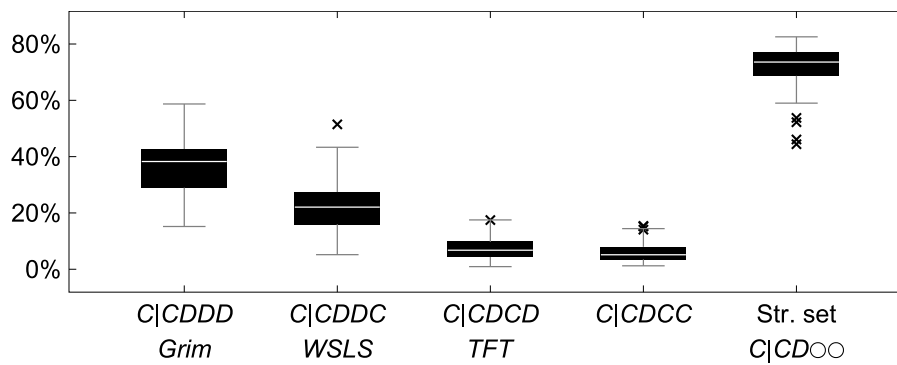


Fig. 3. Box plots of the run-level mean frequencies of the four strategies belonging to the set $C|CD○○$ (where $○$ may be C or D) and their aggregate frequency (rightmost box). Within each run, frequencies were averaged from tick 290 000 to tick 300 000; the box plots summarize 100 independent runs starting from random initial conditions. Outliers are indicated by crosses ($\times$).

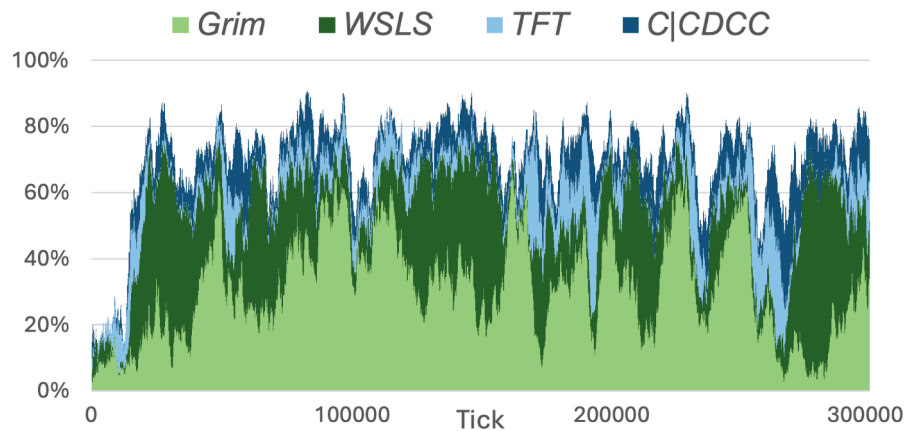


Fig. 4. Percentage of the population using each of the four main strategies, plotted against time, in a representative run of the RPD in the baseline scenario. It is a stacked line graph, so the frequency of *Grim* is the height of the light green line, the frequency of *WSLS* is the distance between the dark green line and the light green line, and so on.

strategy $D|LLLL$. In the absence of action errors, a $D|LLLL$ player in a population of nice strategies obtains the maximal stage payoff $b + 1$ in every interaction. This means that, to be stable, a population must have enough players that defect in the first period. Otherwise the population would be easily invaded by *defect-and-leave*. In particular, *nice* strategies that cooperate with strangers (*TFT*, *WSLS*, *Grim*) can easily be invaded.

Second, after having been exploited (possibly by mistake), players can now leave, rather than stay and defect, and this allows them to escape the potentially long sequence of mutual defections that strategies like *Grim* may suffer after a single mistake. The option to leave creates a different way to discipline defectors: sending them to the pool of singles by breaking the partnership. This punishment is more effective as the proportion of nasty strategies in the pool increases; and we know that there are some nasty strategies in any stable population, because nice strategies do not constitute Nash equilibria in the VRPD under the conditions considered here.

With these two features in mind, it may not come as a surprise that the most successful strategies in the VRPD (Table 2) (1) are *not* nice (97%), (2) continue cooperating after CC (89%), (3) leave after exploitation (82%)—which provides protection against defectors and generates positive assortment—and (4) cooperate after mutual defection (94%). We saw this last feature in *WSLS*, but here it exists for a different reason. In the RPD, *WSLS* cooperates after mutual defection to restore cooperation following an error. Successful strategies in the VRPD cooperate after mutual defection to restore cooperation after the initial defections.

Fig. 5 shows the distribution across runs of the mean frequencies of the three most common strategies in the VRPD in the baseline scenario.

Table 2

Percentage of players using each choice (row) after a given history (column) in the VRPD. Values were calculated as in Table 1. Columns may not sum to 100 because of rounding.

		Empty history	Previous outcome			
			CC	CD	DC	DD
		$\emptyset$				
Choice	C	3	89	4	26	94
	D	97	9	15	46	5
	L		2	82	28	1

We call them *Leavers* because they all respond to exploitation (CD) by leaving. These three strategies are the only strategies that exhibit the four successful traits outlined above. Fig. 5 also includes a box plot for the aggregate frequency of the set of *Leavers* ($D|CL \bullet C$, where $\bullet$ may be C , D , or L) and a final box plot for the aggregate frequency of the set $C|CD○○$ (which includes the successful strategies in the RPD).

Introducing the option to leave changes which strategy set predominates. The set of successful strategies in the RPD ($C|CD○○$) becomes rare when players are given the option to leave (see the rightmost box plot in Fig. 5). It is replaced by the set $D|CL \bullet C$, which accounts for 67.2% of the VRPD population (95% t -confidence interval: [64.3, 70.0]%). By exercising the option to leave, members of this set sustain a higher cooperation rate: at baseline, cooperation rises from less than 40% in the RPD to more than 60% in the VRPD (see Fig. 2).

After its initial defection, *Leaver*₁ ($D|CLDC$) behaves similarly to *WSLS*, except that it responds to a partner's defection by leaving rather

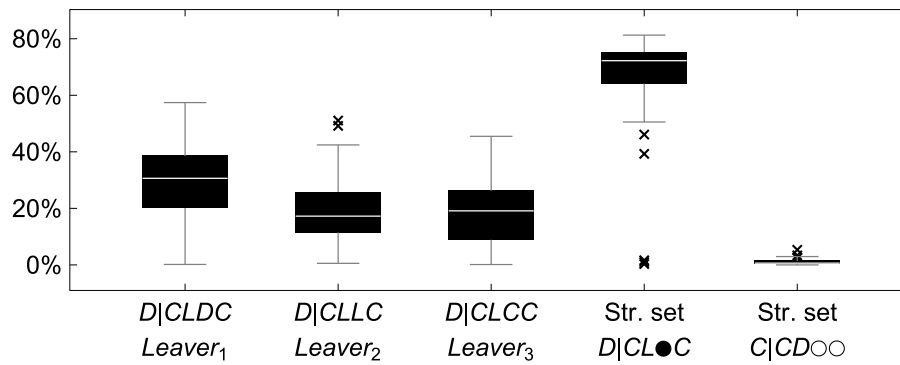


Fig. 5. Box plots of the run-level mean frequencies of the three individual *Leaver* strategies, their aggregate strategy set $D|CL \bullet C$, and the strategy set $C|CD \circ \circ$. Values were calculated and summarized as in Fig. 3. Outliers are indicated by crosses ($\times$). In the definition of the sets, $\bullet$ may be C , D , or L .

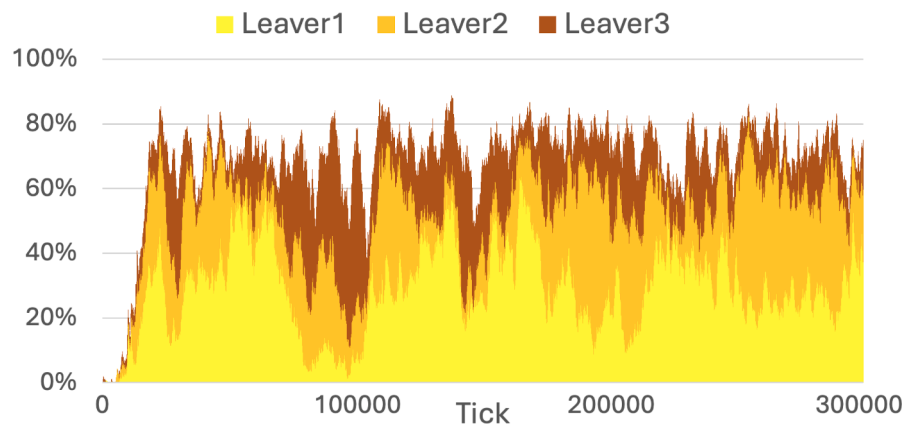


Fig. 6. Percentage of the population using each of the three main strategies plotted against time, in a representative run of the model with the option to leave in the baseline scenario. It is a stacked line graph, so that the frequency of *Leaver*₁ is the height of the yellow line, the frequency of *Leaver*₂ is the distance between the orange line and the yellow line, and so on.

than by defecting. The three *Leavers* differ only in their behavior after exploiting their partner (i.e., after DC). Because 82% of the population leaves after being exploited, this contingency is encountered relatively infrequently and consequently has only a limited effect on fitness. In the absence of errors, any match between two of the three *Leavers* follows the same path: $DD-CC-CC-CC-\dots$. If a single error occurs, however, the partnership is broken.

As in the RPD, in the VRPD there are also strong fluctuations over time and across runs in the frequency of each individual strategy in the set $D|CL \bullet C$ (see Fig. 6). Thus, our main finding concerns the behavior and collective prevalence shared by the three *Leavers*, rather than the differences among their individual frequencies.

Finally, note that, in the VRPD, increasing the expected number of interactions, γ , does not lead to less cooperation, as happens in the RPD (see Fig. 2(b)). The reason is that successful players do not stay and defect after having been exploited; they simply leave and so escape the potentially long sequence of mutual defections that classical strategies may suffer after a single mistake.

Strategies that do not leave suffer a difficult dilemma: if they are not tolerant of errors, like *Grim*, then they do not fare well against themselves; if, on the other hand, they are tolerant of errors, like *WSLS*, then they are vulnerable to exploitation by defectors. Leaving provides a way out of this dilemma: a strategy that performs well against itself can protect its users from defectors by leaving them, and this generates positive assortment, a key principle known to underlie the evolution of cooperation (Eshel and Cavalli-Sforza, 1982; Doebeli and Hauert, 2005; McNamara et al., 2008; Fletcher and Doebeli, 2009). This assortment mechanism is closely related to that identified in models with strategy-

dependent interaction times (Křivan and Cressman, 2017, 2020). In our model, however, differences in partnership duration are generated endogenously by evolving conditional separation responses rather than specified in advance. Under this endogenous assortment mechanism, increasing the expected number of interactions γ allows mutually cooperative partnerships to persist and reap the benefits of cooperation for longer, while individuals remain free to leave exploitative relationships. Consequently, cooperation is substantially higher than in a system where players cannot leave their partners. More freedom at the individual level leads to better outcomes at the population level.

4. Mathematical insights

The mechanisms underlying the simulation results have been examined using analytical work and reduced mean-dynamic systems described below and in the SI. There, we formalize important differences between the RPD and the VRPD in the presence of errors. In the RPD with errors, there are a number of monomorphic strict Nash states to which many evolutionary dynamics can converge, but that is not the case in the VRPD. For instance, in our baseline scenario, *Grim*, *WSLS*, *AllD*, and $D|CDDC$ (*nWSLS*, for *nasty WSLS*) are strict Nash in the RPD, but they are not even evolutionarily stable in the VRPD.

In fact, focusing on the donation game, we show that no memory-one strategy is evolutionarily stable (or strict Nash) in the VRPD for small to moderate error rates ($\epsilon \in (0, 0.1)$) and $\delta \in [0.5, 1)$, even when potential mutants are restricted to the class of memory-one strategies. An evolutionarily stable strategy is one that, when common in the population, cannot be invaded by any rare mutant. The intuition behind

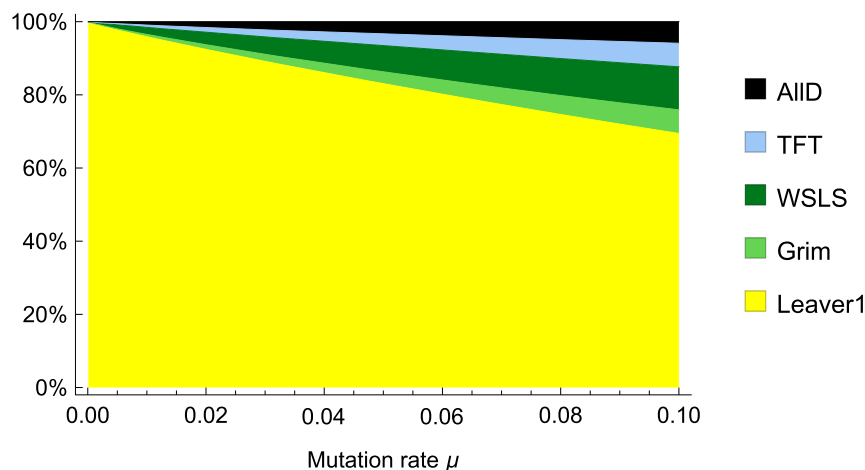


Fig. 7. Strategy distribution of the attracting rest point found numerically for different mutation rates. Other parameter values are set as in the baseline scenario. This rest point is not necessarily stable if the strategy set is enlarged.

our result is that endogenous separation creates behavioral redundancies among memory-one strategies. If a strategy leaves after a particular outcome, some prescriptions governing how play would subsequently continue within that partnership become behaviorally irrelevant. Conversely, if both players' responses after an outcome reproduce the actions they would take at the beginning of a new partnership, continuing together can be outcome-equivalent to separating and being rematched. For 156 of the 162 memory-one strategies, these redundancies allow us to construct a distinct neutral mutant that generates the same distribution of outcomes, and hence the same payoffs, in every resident-mutant pairing. The resident therefore fails the conditions for evolutionary stability. The remaining six strategies do not admit this neutral-mutant construction, but each can be invaded by a mutant that earns a strictly higher payoff under the parameter conditions stated above. The formal construction and the treatment of the six exceptional strategies are provided in the SI.

In the SI, we also derive the mean dynamic (MD) of our evolutionary process. The MD (Sandholm, 2010) is a deterministic system of ODEs that approximates the stochastic population dynamics over finite time spans for a fixed strategy set. Because evaluating the endogenous payoff functions throughout the state space becomes computationally prohibitive when all 162 memory-one strategies are included, we analyze reduced systems containing a small number of strategies. These reduced systems are not intended to reproduce the precise long-run strategy distribution of the full stochastic simulations. Rather, they allow us to investigate attractors and isolate the mechanisms underlying the success of strategies that exercise the option to leave.

In particular, we consider a five-strategy system comprising *Leaver₁* and four strategies that are prevalent or theoretically important in the RPD: *Grim*, *WSLS*, *TFT*, and *AIID*. With the baseline parameters, trajectories starting from 1000 randomly selected initial states all converge to the same rest point, at which more than 80% of the population uses *Leaver₁*. This result is qualitatively consistent with the full simulations, in which the three *Leaver* strategies collectively predominate in the VRPD, whereas the conventional non-leaving strategies that are successful in the RPD become uncommon (Fig. 5). Naturally, this comparison concerns the qualitative evolutionary pattern rather than the precise strategy frequencies. The latter are not expected to coincide because the full simulations admit all 162 strategies and use local mutation, whereas the reduced MD contains only five strategies and assumes uniform experimentation among them.

Fig. 7 shows how the composition of the attracting rest point found numerically changes with the mutation rate μ . As μ approaches zero, the frequency of *Leaver₁* approaches one, while the frequencies of the four non-leaving strategies approach zero.

We also examine the attractors of this reduced system for different error rates (see Fig. 8). For each value of ϵ , we numerically solve the MD from each of the five monomorphic states. For $\epsilon \leq 10^{-3}$, all five trajectories converge to the same rest point, where more than 75% of the population uses *TFT*, *WSLS*, or *Grim*. Within the thin red interval in Fig. 8, the trajectories converge to two different rest points. For $\epsilon \geq 2 \times 10^{-3}$, all five trajectories converge to a qualitatively different rest point characterized by the predominance of *Leaver₁*.

Fig. 8 shows that the dynamics at a fixed positive but small error rate, such as $\epsilon = 2 \times 10^{-3}$, can differ qualitatively from those obtained at $\epsilon = 0$ or in the vanishing-error limit $\epsilon \rightarrow 0$. Consequently, conclusions derived under zero error or by taking the error rate to zero need not apply when the error rate is small but fixed and positive. This distinction is examined further in the SI, Section "Fixed positive error rates and the vanishing-error limit."

5. Leaving rather than punishing by defection: positive assortment

A comparison of *Leaver₁* (*D|CLDC*) and *nWSLS* (*D|CDDC*) illustrates why leaving unilateral defectors (as *Leaver₁* does) is better than punishing them by defecting (like *nWSLS*). First, notice that in a population made up of any mixture of *Leaver₁* and *nWSLS* both strategies obtain the same expected payoff (see Fig. 9(a)) because in every partnership individuals expect the same sequence of outcomes. Without errors, every partnership follows the sequence *DD-CC-CC-...*. Following a unilateral error, play returns, in the absence of further errors, to the sequence *DD-CC-CC-...*: *Leaver₁* achieves this through separation and rematching, whereas *nWSLS* does so within the existing partnership. Fig. 9(a) includes the (1-dimensional) mean dynamic at the bottom: arrows point in the direction of movement and circles indicate the location of the asymptotically stable rest points.

Next, consider the fitness for *nWSLS* and *Leaver₁* interacting with *AIID* (see Figs. 9(b) and 9(c) respectively). An individual with either strategy obtains the same payoff in a population of *AIID*, and a single *AIID* obtains the same payoff against either of them when entering a monomorphic population. In both cases a player using *AIID* can expect an alternating sequence of *DD* and *DC* outcomes (either with the same partner if in a *nWSLS* population, or changing partners after *DC* if in a *Leaver₁* population).

However, the fitnesses of *nWSLS* and *Leaver₁* differ in populations with substantial frequency of *AIID* as shown in Fig. 9. Panel 9(a) shows that both strategies are equivalent in an isolated bimorphic population. Panel 9(b) shows the dynamics in a population of *nWSLS* and *AIID*: linear payoff functions and a single attractor where most players use

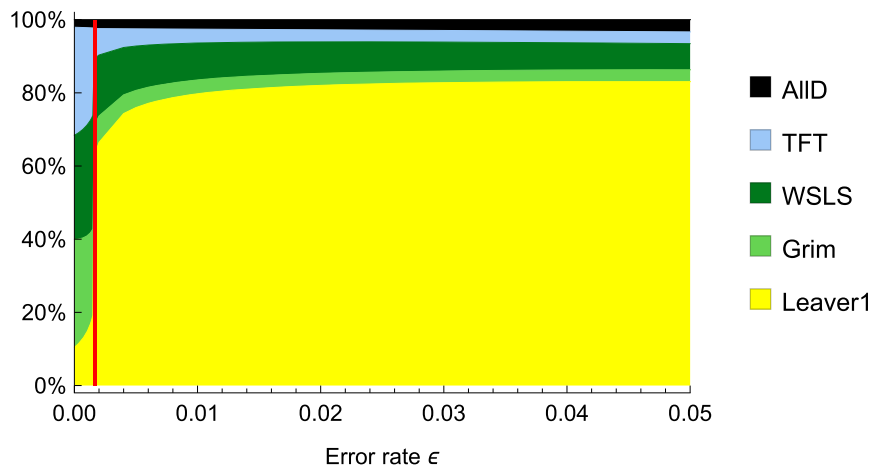


Fig. 8. Strategy distributions reached by numerical solutions of the MD for different values of the error rate $\epsilon \in [0, 0.05]$. For each value of ϵ , the MD is numerically solved starting from each of the five corners of the simplex. Within the red interval, trajectories converge to two different rest points; outside this interval, all five trajectories converge to the distribution shown. Other parameter values are set as in the baseline scenario.

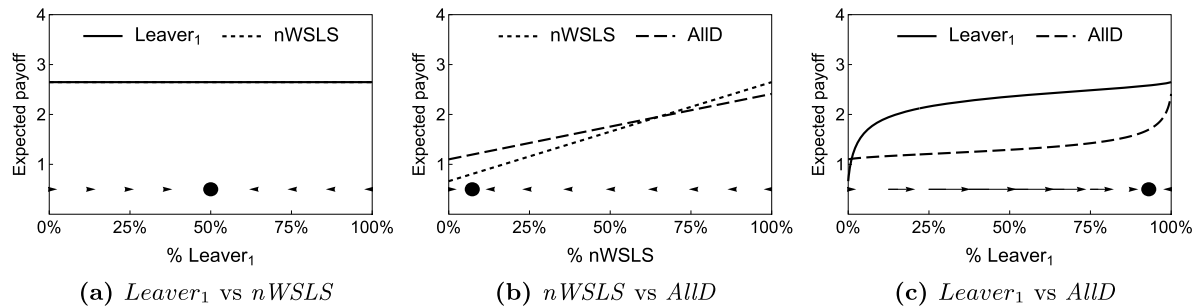


Fig. 9. Payoff functions and phase portraits for several bimorphic populations. Parameters are set as in the baseline scenario ($\delta = 0.95$ and $\epsilon = \mu = 0.05$). Mutation in these figures consists in choosing one of the two strategies at random. Note that errors affect the expected payoff obtained by each strategy, but mutation does not.

Table 3

Assortment and payoffs in a bimorphic population with 75% of *AIID* players and 25% of *nWSLS* players. $\delta = 0.95$ and $\epsilon = 0.05$.

	Fraction matched to		Payoff with		Payoff
	<i>nWSLS</i> (25%)	<i>AIID</i> (75%)	<i>nWSLS</i>	<i>AIID</i>	
<i>nWSLS</i>	25%	75%	2.65	0.66	1.16
<i>AIID</i>	25%	75%	2.41	1.10	1.43

Table 4

Assortment and payoffs in a bimorphic population with 75% of *AIID* players and 25% of *Leaver1* players. $\delta = 0.95$ and $\epsilon = 0.05$.

	Fraction matched to		Payoff with		Payoff
	<i>Leaver1</i> (25%)	<i>AIID</i> (75%)	<i>Leaver1</i>	<i>AIID</i>	
<i>Leaver1</i>	75.37%	24.63%	2.65	0.66	2.16
<i>AIID</i>	8.21%	91.79%	2.41	1.10	1.21

AIID. Panel 9(c) shows the dynamics in a population of *Leaver1* and *AIID*: nonlinear payoff functions (with the same initial and final values as in panel 9(b)) and a single attractor where most players use *Leaver1*.

To understand why *nWSLS* and *Leaver1* differ, assume that the frequency of *AIID* defectors is 75%. The first part of Table 3 presents the expected fraction of *nWSLS* players in partnership with *nWSLS* or with *AIID* players. Since partnerships are never endogenously broken, those are simply the proportions of each strategy in the population, and the same holds for *AIID* players. The second part of the table shows the average payoff obtained by *nWSLS* and *AIID* players in partnerships with either *nWSLS* or *AIID*. The last column shows the average payoff obtained by each strategy in the population. Since the payoff to *AIID* (1.43) is higher than the payoff to *nWSLS* (1.16), the proportion of *AIID* will increase (see panel 9(b)).

Table 4 presents the corresponding data for *Leaver1* and *AIID*. In this case, homogeneous partnerships in which both players use the same strategy are stable, while heterogeneous partnerships separate after playing together twice (unless there is an error). As a result, most

Leaver1 players are found in partnerships with other *Leaver1* players, and most *AIID* players are found in partnerships with other *AIID* players.

The option to leave generates positive assortment: the fraction of *Leaver1* players who are in partnerships with other *Leaver1* players is larger than the proportion of *Leaver1* players in the population; and the same holds for *AIID* players. As a result of positive assortment, the average payoff to *Leaver1* players in the population is close to the payoff obtained when two such players interact (this effect pushes its payoff function upwards in Fig. 9(c) compared to 9(b)); and the same holds for *AIID* players (which pushes its payoff function downwards in Fig. 9(c)).

In the situation shown in Table 4, the proportion of *AIID* will decrease, since the payoff to *AIID* (1.21) is lower than the payoff to *Leaver1* (2.16) (see panel 9(c)).

6. Conclusions

We have studied population dynamics in the repeated Prisoner's Dilemma and in the voluntarily repeated Prisoner's Dilemma with action errors and recurrent exploration of alternative strategies. Both fea-

tures are commonly observed in real-life interactions but have often been considered separately. Studying them together leads to insights not captured by previous models.

In settings with non-negligible error and mutation rates, a small set of strategies typically predominates. Although the option to leave precludes a fully cooperative equilibrium, it substantially increases cooperation when relationships persist long enough for contingent behavior to matter (see also Wubs et al., 2016). This increase in cooperation becomes larger as the expected number of interactions rises.

The predominant strategy set is completely different with the option to leave. Strategies that exercise the option to leave are more robust to errors and thrive compared to strategies that never leave. With errors, non-leaving strategies such as *Tit-for-Tat*, *Win-Stay-Lose-Shift* and *Grim* face a dilemma: forgiveness invites exploitation, while strictness causes long periods of miscoordination. Leaving resolves this by letting players abandon exploiters while still cooperating with similar types.

As a result, strategies that leave unilateral defectors outperform those that punish by defecting within partnerships, and cooperation increases markedly. Importantly, the option to leave not only provides an effective alternative to in-partnership retaliation, it also generates positive assortment, allowing cooperative strategies that leave defectors to interact more often with each other. Cooperative partnerships persist, while exploiters often return to the pool of singles, which is dominated by defectors. Thus, most cooperators are found in partnerships with other cooperators, and most exploiters are found in partnerships with other exploiters. This allows for the evolutionary emergence and maintenance of high levels of cooperation.

In sum, our results highlight the central role of the option to leave in supporting (partial) cooperation under realistic conditions. Giving agents the freedom to leave exploitative relationships can lead to better outcomes at the population level.

Materials and methods

The analytical framework used in this study is explained in detail in the Supplementary Information (SI), which also includes the mathematical analysis and proofs. All software developed for this work has been released under the GNU GPL v3 license and is publicly available (Izquierdo et al., 2026).

CRedit authorship contribution statement

Luis R. Izquierdo: Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization; **Segismundo S. Izquierdo:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization; **Robert Boyd:** Writing – review & editing, Writing – original draft, Software, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work, the authors used ChatGPT and Codex (OpenAI) to assist with language editing and improve the clarity and organization of the manuscript. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Funding

This work was supported by the Spanish State Research Agency (grant numbers PID2024-159461NB-I00/MICIU and PID2020-118906GB-I00/MCIN), the Junta de Castilla y León and EU-FEDER program (grant number CLU-2025-2-05 - BIOECOUVA), and the

Institute of Human Origins, Arizona State University. Open access funding provided by Universidad de Burgos.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank two anonymous reviewers for their careful reading and constructive comments, which helped us improve the manuscript.

Data and code availability

All simulation data reported in this article can be generated using the software developed for this work, which has been released under the GNU GPL v3 license and is publicly available (Izquierdo et al., 2026).

Supplementary material

Supplementary material associated with this article can be found in the online version at [10.1016/j.jtbi.2026.112598](https://doi.org/10.1016/j.jtbi.2026.112598).

References

- Aktipis, C.A., 2004. Know when to walk away: contingent movement and the evolution of cooperation. *J. Theor. Biol.* 231, 249–260. <https://doi.org/10.1016/j.jtbi.2004.06.020>
- Aktipis, C.A., 2011. Is cooperation viable in mobile organisms? Simple walk away rule favors the evolution of cooperation in groups. *Evol. Hum. Behav.* 32, 263–276. <https://doi.org/10.1016/J.EVOLHUMBHAV.2011.01.002>
- Axelrod, R., 1984. *The Evolution of Cooperation*. Basic Books.
- Barclay, P., 2013. Strategies for cooperation in biological markets, especially for humans. *Evol. Hum. Behav.* 34, 164–175. <https://doi.org/10.1016/j.evolhumbehav.2013.02.002>
- Barclay, P., Raihani, N., 2016. Partner choice versus punishment in human prisoner's dilemmas. *Evol. Hum. Behav.* 37 (4), 263–271. <https://doi.org/10.1016/j.evolhumbehav.2015.12.004>
- Bendor, J., Swistak, P., 1995. Types of evolutionary stability and the problem of cooperation. *Proc. Natl. Acad. Sci. U. S. A.* 92 (8), 3596–3600. <https://doi.org/10.1073/pnas.92.8.3596>
- Bendor, J., Swistak, P., 1998. Evolutionary equilibria: characterization theorems and their implications. *Theory Decis.* 45, 99–159. <https://doi.org/10.1023/A:1005083323183>
- Boyd, R., 1989. Mistakes allow evolutionary stability in the repeated prisoner's dilemma game. *J. Theor. Biol.* 136, 47–56. [https://doi.org/10.1016/S0022-5193\(89\)80188-2](https://doi.org/10.1016/S0022-5193(89)80188-2)
- Boyd, R., Lorberbaum, J.P., 1987. No pure strategy is evolutionarily stable in the repeated prisoner's dilemma game. *Nature* 327 (6117), 58–59. <https://doi.org/10.1038/327058a0>
- Boyd, R., Mathew, S., 2021. Arbitration supports reciprocity when there are frequent perception errors. *Nat. Hum. Behav.* 5 (5), 596–603. <https://doi.org/10.1038/s41562-020-01008-1>
- Broom, M., Křivan, V., 2020. Two-strategy games with time constraints on regular graphs. *J. Theor. Biol.* 506, 110426. <https://doi.org/10.1016/j.jtbi.2020.110426>
- Chabris, C.F., Lee, J.J., Cesarini, D., Benjamin, D.J., Laibson, D.I., 2015. The fourth law of behavior genetics. *Curr. Dir. Psychol. Sci.* 24 (4), 304–312. <https://doi.org/10.1177/0963721415580430>
- Doebeli, M., Hauert, C., 2005. Models of cooperation based on the prisoner's dilemma and the snowdrift game. *Ecol. Lett.* 8 (7), 748–766. <https://doi.org/10.1111/j.1461-0248.2005.00773.x>
- Enquist, M., Leimar, O., 1993. The evolution of cooperation in mobile organisms. *Anim. Behav.* 45, 747–757. <https://doi.org/10.1006/anbe.1993.1089>
- Eshel, I., Cavalli-Sforza, L.L., 1982. Assortment of encounters and evolution of cooperativeness. *Proc. Natl. Acad. Sci.* 79 (4), 1331–1335. <https://doi.org/10.1073/pnas.79.4.1331>
- Fehl, K., van der Post, D.J., Semmann, D., 2011. Co-evolution of behaviour and social network structure promotes human cooperation. *Ecol. Lett.* 14, 546–551. <https://doi.org/10.1111/j.1461-0248.2011.01615.x>
- Fletcher, J.A., Doebeli, M., 2009. A simple and general explanation for the evolution of altruism. *Proc. R. Soc. B: Biol. Sci.* 276 (1654), 13–19. <https://doi.org/10.1098/rspb.2008.0829>
- Friedman, J.W., Hammerstein, P., 1991. To trade, or not to trade; that is the question. In: Selten, R. (Eds.), *Game Equilibrium Models I*. Springer, Berlin, Heidelberg, pp. 257–275. https://doi.org/10.1007/978-3-662-02674-8_9
- Fudenberg, D., Imhof, L.A., 2006. Imitation processes with small mutations. *J. Econ. Theory* 131, 251–262. <https://doi.org/10.1016/j.jet.2005.04.006>
- Fudenberg, D., Nowak, M.A., Taylor, C., Imhof, L.A., 2006. Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theor. Popul. Biol.* 70, 352–363. <https://doi.org/10.1016/j.tpb.2006.07.006>

- Fujiwara-Greve, T., Okuno-Fujiwara, M., 2009. Voluntarily separable repeated prisoner's dilemma. *Rev. Econ. Stud.* 76 (3), 993–1021. <https://doi.org/10.1111/j.1467-937X.2009.00539.x>
- García, J., van Veelen, M., 2016. In and out of equilibrium I: evolution of strategies in repeated games with discounting. *J. Econ. Theory* 161, 161–189. <https://doi.org/10.1016/j.jet.2015.11.007>
- Graser, C., Fujiwara-Greve, T., García, J., van Veelen, M., 2025. Repeated games with partner choice. *PLOS Comput. Biol.* 21, e1012810. <https://doi.org/10.1371/journal.pcbi.1012810>
- Grujić, J., Gracia-Lázaro, C., Milinski, M., Semmann, D., Traulsen, A., Cuesta, J.A., Moreno, Y., Sánchez, A., 2014. A comparative analysis of spatial prisoner's dilemma experiments: conditional cooperation and payoff irrelevance. *Sci. Rep.* 4, 4615. <https://doi.org/10.1038/srep04615>
- Gutiérrez-Mielgo, A., Izquierdo, L.R., Izquierdo, S.S., 2025. Games with costly endogenous separation. In: *Proceedings of the 1st International Electronic Conference on Games*. <https://uvadoc.uva.es/handle/10324/80280>
- Hamilton, I.M., Taborsky, M., 2005. Contingent movement and cooperation evolve under generalized reciprocity. *Proc. R. Soc. B: Biol. Sci.* 272 (1578), 2259–2267. <https://doi.org/10.1098/rspb.2005.3248>
- Hayashi, N., Yamagishi, T., 1998. Selective play: choosing partners in an uncertain world. *Personal. Soc. Psychol. Rev.* 2, 276–289. https://doi.org/10.1207/s15327957pspr0204_4
- Hilbe, C., Martínez-Vaquero, L.A., Chatterjee, K., Nowak, M.A., 2017. Memory-n strategies of direct reciprocity. *Proc. Natl. Acad. Sci.* 114 (18), 4715–4720. <https://doi.org/10.1073/pnas.1621239114>
- Hilbe, C., Nowak, M.A., Sigmund, K., 2013. Evolution of extortion in iterated prisoner's dilemma games. *Proc. Natl. Acad. Sci.* 110 (17), 6913–6918. <https://doi.org/10.1073/pnas.1214834110>
- Imhof, L.A., Fudenberg, D., Nowak, M.A., 2005. Evolutionary cycles of cooperation and defection. *Proc. Natl. Acad. Sci.* 102 (31), 10797–10800. <https://doi.org/10.1073/pnas.0502589102>
- Izquierdo, L.R., Izquierdo, S.S., Boyd, R., 2026. ES-1-outcome. Software. <https://doi.org/10.5281/zenodo.17473910>
- Izquierdo, S.S., Izquierdo, L.R., 2026. Stable strategies in repeated games with endogenous separation. *Dyn. Games Appl.* In Press.
- Izquierdo, S.S., Izquierdo, L.R., Vega-Redondo, F., 2010. The option to leave: conditional dissociation in the evolution of cooperation. *J. Theor. Biol.* 267 (1), 76–84. <https://doi.org/10.1016/j.jtbi.2010.07.039>
- Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29–56. <https://doi.org/10.2307/2951777>
- Kim, Y.G., 1994. Evolutionarily stable strategies in the repeated prisoner's dilemma. *Math. Soc. Sci.* 28, 167–197. [https://doi.org/10.1016/0165-4896\(94\)90002-7](https://doi.org/10.1016/0165-4896(94)90002-7)
- Kokko, H., Brooks, R., McNamara, J.M., Houston, A.I., 2002. The sexual selection continuum. *Proc. R. Soc. Lond. B Biol. Sci.* 269 (1498), 1331–1340. <https://doi.org/10.1098/rspb.2002.2020>
- Kraines, D., Kraines, V., 1993. Learning to cooperate with Pavlov: an adaptive strategy for the iterated prisoner's dilemma with noise. *Theory Decis.* 35 (2), 107–150. <https://doi.org/10.1007/BF01074955>
- Křivan, V., Cressman, R., 2017. Interaction times change evolutionary outcomes: two-player matrix games. *J. Theor. Biol.* 416, 199–207. <https://doi.org/10.1016/j.jtbi.2017.01.010>
- Křivan, V., Cressman, R., 2020. Defectors' intolerance of others promotes cooperation in the repeated public goods game with opting out. *Sci. Rep.* 10, 19511. <https://doi.org/10.1038/s41598-020-76506-3>
- Kurokawa, S., 2019. Three-player repeated games with an opt-out option. *J. Theor. Biol.* 480, 13–22. <https://doi.org/10.1016/j.jtbi.2019.07.012>
- Leimar, O., 1997. Repeated games: a state space approach. *J. Theor. Biol.* 184 (4), 471–498. <https://doi.org/10.1006/jtbi.1996.0286>
- Mathew, S., Boyd, R., 2026. Rethinking reciprocity. *Trends Cogn. Sci.* . <https://doi.org/10.1016/j.tics.2026.06.007>
- McNamara, J.M., Barta, Z., Fromhage, L., Houston, A.I., 2008. The coevolution of choosiness and cooperation. *Nature* 451, 189–192. <https://doi.org/10.1038/nature06455>
- McNamara, J.M., Barta, Z., Houston, A.I., 2004. Variation in behaviour promotes cooperation in the prisoner's dilemma game. *Nature* 428 (6984), 745–748. <https://doi.org/10.1038/nature02432>
- McNamara, J.M., Leimar, O., 2010. Variation and the response to variation as a basis for successful cooperation. *Philos. Trans. R. Soc. B: Biol. Sci.* 365 (1553), 2627–2633. <https://doi.org/10.1098/rstb.2010.0159>
- Noë, R., Hammerstein, P., 1995. Biological markets. *Trends Ecol. Evol.* 10, 336–339. [https://doi.org/10.1016/S0169-5347\(00\)89123-5](https://doi.org/10.1016/S0169-5347(00)89123-5)
- Nowak, M., Sigmund, K., 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* 364 (6432), 56–58. <https://doi.org/10.1038/364056a0>
- Rand, D.G., Arbesman, S., Christakis, N.A., 2011. Dynamic social networks promote cooperation in experiments with humans. *Proc. Natl. Acad. Sci.* 108, 19193–19198. <https://doi.org/10.1073/pnas.1108243108>
- Sandholm, W.H., 2010. *Population games and evolutionary dynamics*. The MIT Press.
- Schmid, L., Chatterjee, K., Hilbe, C., Nowak, M.A., 2021. A unified framework of direct and indirect reciprocity. *Nat. Hum. Behav.* 5 (10), 1292–1302. <https://doi.org/10.1038/s41562-021-01114-8>
- Stewart, A.J., Parsons, T.L., Plotkin, J.B., 2016. Evolutionary consequences of behavioral diversity. *Proc. Natl. Acad. Sci.* 113 (45), E7003–E7009. <https://doi.org/10.1073/pnas.1608990113>
- Stewart, A.J., Plotkin, J.B., 2013. From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proc. Natl. Acad. Sci.* 110 (38), 15348–15353. <https://doi.org/10.1073/pnas.1306246110>
- Sugden, R., 1986. *The economics of rights, co-operation and welfare*. Basil Blackwell, Oxford.
- Tkadle, J., Hilbe, C., Nowak, M.A., 2023. Mutation enhances cooperation in direct reciprocity. *Proc. Natl. Acad. Sci.* 120, e2221080120. <https://doi.org/10.1073/pnas.2221080120>
- Traulsen, A., Hauert, C., 2009. Stochastic evolutionary game dynamics. In: Schuster, H.G. (Ed.), *Reviews of Nonlinear Dynamics and Complexity*. Wiley, New York. Vol. 2, pp. 25–61. <https://doi.org/10.1002/9783527628001.ch2>
- Traulsen, A., Semmann, D., Sommerfeld, R.D., Krambeck, H.-J., Milinski, M., 2010. Human strategy updating in evolutionary games. *Proc. Natl. Acad. Sci.* 107, 2962–2966. <https://doi.org/10.1073/pnas.0912515107>
- van Veelen, M., García, J., 2019. In and out of equilibrium II: evolution in repeated games with discounting and complexity costs. *Games Econ. Behav.* 115, 113–130. <https://doi.org/10.1016/j.geb.2019.02.013>
- Wu, B., Gokhale, C.S., Wang, L., Traulsen, A., 2012. How small are small mutation rates? *J. Math. Biol.* 64 (5), 803–827. <https://doi.org/10.1007/s00285-011-0430-8>
- Wubs, M., Bshary, R., Lehmann, L., 2016. Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions. *Proc. R. Soc. B: Biol. Sci.* 283 (1832), 20160488. <https://doi.org/10.1098/rspb.2016.0488>
- Young, H.P., 1993. The evolution of conventions. *Econometrica* 61, 57–84. <https://doi.org/10.2307/2951778>
- Young, H.P., 1998. *Individual Strategy and Social Structure*. Princeton University Press, Princeton.
- Zhang, B.-Y., Fan, S.-J., Li, C., Zheng, X.-D., Bao, J.-Z., Cressman, R., Tao, Y., 2016. Opting out against defection leads to stable coexistence with cooperation. *Sci. Rep.* 6, 35902. <https://doi.org/10.1038/srep35902>