



# Evolution of noisy learning in games

Marta C. Couto<sup>a,b,1</sup> , Fernando P. Santos<sup>a</sup>, and Christian Hilbe<sup>b,c</sup>

Edited by Marcus W. Feldman, Stanford University, Stanford, CA; received October 23, 2025; accepted April 21, 2026

People make strategic decisions many times a day—during negotiations, when coordinating actions with others, or when choosing partners for cooperation. The resulting dynamics can be studied with learning theory and evolutionary game theory. These frameworks explore how people adapt their decisions over time, in light of how effective their strategies have been. The outcomes of such learning processes depend on how sensitive individuals are to the performance of their strategies. When they are more sensitive, they systematically favor strategies they deem more successful. When they are less sensitive, their learning process is noisier and more erratic. Traditionally, most models treat this sensitivity as a fixed parameter—like the “selection strength” parameter in evolutionary models. Instead, we study how strategies and sensitivities coevolve. We find that the coevolutionary endpoints depend on both the type of strategic interaction and the learning rule employed. In prisoner’s dilemmas, we often observe sensitivities to increase indefinitely. But in snowdrift and stag-hunt games, sensitivities often converge to a finite value, or we observe evolutionary branching altogether. These results shed light on how evolution might shape learning mechanisms for social behavior. They suggest that noisy learning does not need to be a by-product of cognitive constraints. Instead, it can serve as a means to gain strategic advantages.

evolutionary game theory | adaptive dynamics | learning | social dilemma

People constantly make strategic decisions. These decisions range from inconsequential choices (e.g., whether to do a small favor) to major commitments (e.g., choosing a business partner). The process by which individuals arrive at their decisions can be analyzed with learning theory (1–3) or evolutionary game theory (4–9). To this end, individuals are regarded as players who interact in games. They can choose among different strategies and win a payoff that depends both on their own and their coplayers’ decisions. Importantly, they do not need to behave optimally from the outset. Rather, individuals are assumed to engage in the same interaction (or sufficiently similar ones) many times. Over the course of these interactions, they can adapt their behavior to yield better payoffs.

The outcome of this adaptation process depends on how sensitive players are to their strategies’ performance. Most models capture this sensitivity by a parameter. For example, in stochastic learning models of evolutionary game theory, outcomes crucially depend on the selection strength or intensity of selection,  $\beta$ . This parameter  $\beta$  determines to which extent individuals switch their strategies because of payoff differences (Fig. 1A). When this parameter is small, players are largely insensitive to their payoffs. They mostly switch from one strategy to another by chance (Fig. 1B, Left). The respective limit of weak selection ( $\beta \rightarrow 0$ ) has become a major paradigm in evolutionary game theory, partly because it simplifies many calculations (22–26). In this limit, learning takes the form of an (almost unbiased) random walk on the space of all strategies. In contrast, as selection strength becomes large, players only switch strategies if the new strategy is superior (Fig. 1B, Right). The respective limit of strong selection ( $\beta \rightarrow \infty$ ) has become another major paradigm to study evolutionary games (27–29). Respective results are closely aligned with key concepts in classical game theory. For example, in this limit, strict Nash equilibria correspond to absorbing states of the learning process (30, 31). In the following, we refer to  $\beta$  as a player’s (payoff) sensitivity. A small  $\beta$  translates into a more noisy learning process. A large  $\beta$  makes individuals update their strategies more deterministically. Related parameters also exist in other learning models. For example, in experience-weighted attraction learning (32–34), outcomes critically depend on a player’s sensitivity to attractions.

Except for a handful of studies discussed further below (35, 36), most models interpret a player’s (payoff) sensitivity as a fixed quantity, often assumed to be the same for everyone. Here, we relax this assumption. We address two related questions. First, we are interested in learning processes among individuals with fixed but different sensitivities. We refer to individuals with larger  $\beta$  as more sensitive learners, and those with lower  $\beta$  as less

## Significance

In strategic interactions, people can improve their performance by reasoning about their available strategies. This reasoning process can be captured with models of learning and evolutionary game theory. These models often contain a parameter that reflects how likely individuals are to switch to strategies they deem more attractive. This parameter has been referred to as an individual’s sensitivity, or as the strength of selection. The smaller this parameter, the noisier the learning process becomes. Herein, we study how this sensitivity itself may evolve over time. We find many scenarios where this sensitivity increases indefinitely. However, we also identify situations where it converges toward a finite value. These results help us understand how noisy strategy updating may result in long-term advantages.

Author affiliations: <sup>a</sup>Socially Intelligent Artificial Systems Group, Informatics Institute, University of Amsterdam, Amsterdam 1098XH, The Netherlands; <sup>b</sup>Max Planck Research Group on the Dynamics of Social Behavior, Max Planck Institute for Evolutionary Biology, Plön 24306, Germany; and <sup>c</sup>Interdisciplinary Transformation University, Linz 4040, Austria

Author contributions: M.C.C., F.P.S., and C.H. designed research; M.C.C. and C.H. performed research; M.C.C. analyzed data; and M.C.C., F.P.S., and C.H. wrote the paper.

The authors declare no competing interest.

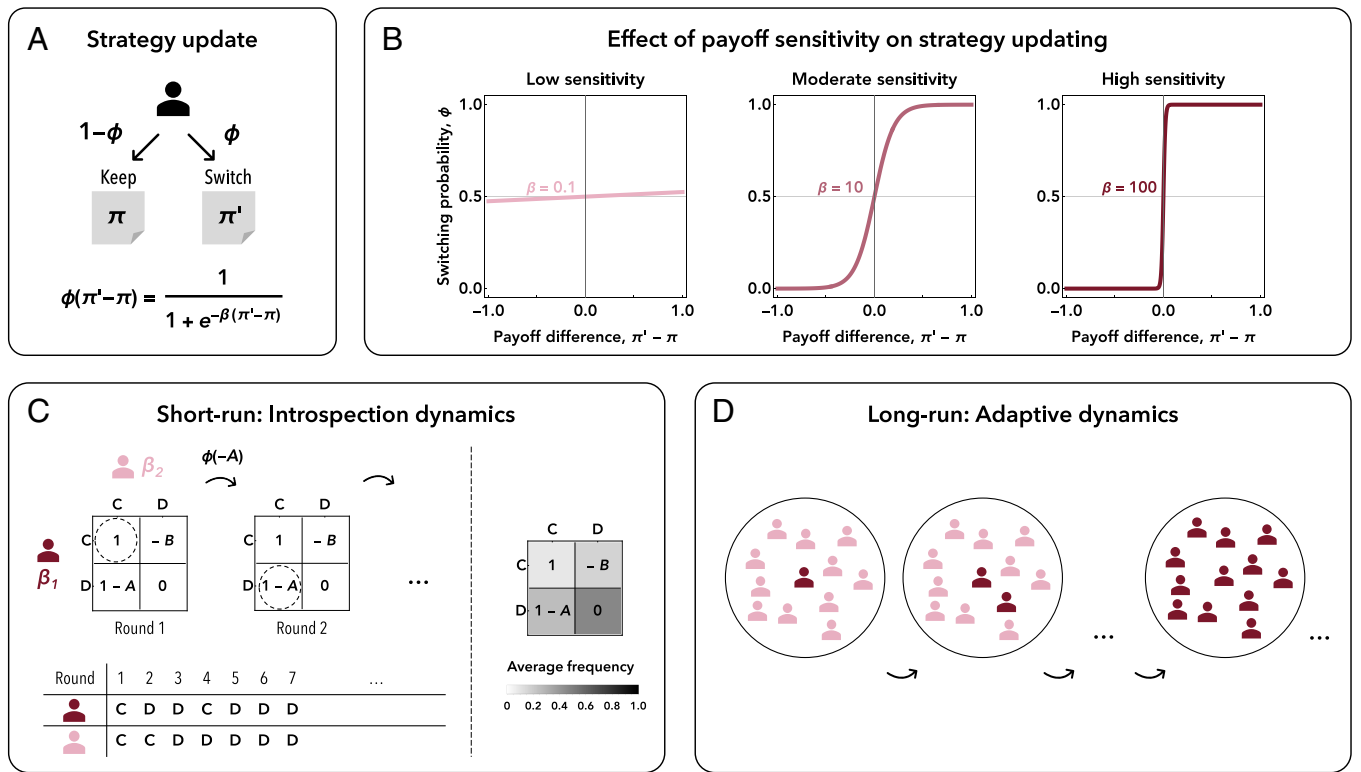
This article is a PNAS Direct Submission.

Copyright © 2026 the Author(s). Published by PNAS. This open access article is distributed under Creative Commons Attribution License 4.0 (CC BY).

<sup>1</sup>To whom correspondence may be addressed. Email: mccouto@evolbio.mpg.de or m.gomesdacunhacouto@uva.nl.

This article contains supporting information online at <https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2529959123/-/DCSupplemental>.

Published May 12, 2026.



**Fig. 1.** An overview of the model. (A) In stochastic models of evolutionary game theory, individuals continually get a chance to revise their strategy. We model this revision process with introspection dynamics (10–17). Here, a player compares its current strategy to a randomly chosen alternative. If  $\pi$  is the payoff of the current strategy, and  $\pi'$  the payoff of the alternative, the player's switching probability is given by a Fermi function  $\phi(\pi' - \pi)$ . This implies players are more likely to switch if the alternative strategy is more beneficial. (B) The exact shape of the switching probability depends on a parameter  $\beta$ . This parameter is often referred to as the strength of selection. Because here we interpret  $\beta$  as an individual (and evolvable) trait, we refer to it as the player's *payoff sensitivity*. For low  $\beta$ , the player's learning process is more noisy. Here, strategy changes are mostly driven by chance. As  $\beta$  becomes larger, updating decisions become more deterministic. Here, individuals increasingly favor those alternative strategies with high payoffs. We are interested in the evolution of this parameter  $\beta$ . Our process unfolds on two time scales. (C) In the short run, each player's payoff sensitivity  $\beta$  is fixed. Given their sensitivity, they choose between different strategies in a stage game. The depicted example shows a round in which player 1 switches from strategy C (cooperation) to strategy D (defection). We iterate this learning process for many rounds. Based on these iterations, we compute the average frequency of each possible game outcome (illustrated as a black-and-white gradient). This allows us to compute the players' expected payoff as a function of their  $\beta$  values. (D) In the long run, we let the players' payoff sensitivity  $\beta$  evolve. We model this long-run process with adaptive dynamics (18–21).

sensitive learners. We ask to what extent more sensitive learners are able to systematically outperform their opponents.

Second, we explore the dynamics when a player's sensitivity is itself an evolving trait. To this end, we consider a framework akin to the indirect evolutionary approach used to study the evolution of preferences (37–41). The dynamics unfolds on two time scales. In the short-run, sensitivities are fixed. Players interact in a given game with others, adopt new strategies based on their sensitivity, and as a result, earn a certain average payoff (Fig. 1C). In the long-run, we allow their sensitivity to evolve, based on the average payoffs they yield. To the extent that a player's sensitivity is an inheritable trait, this assumption may reflect that more effective learners have more offspring (Fig. 1D). We explore this long-run dynamics analytically and with simulations, based on the adaptive dynamics formalism (18–21).

To get an intuitive understanding, we focus on the simplest class of games: symmetric games among two players with two strategies. These games include the prisoner's dilemma, the stag-hunt game, and the snowdrift game. However, we also show that the respective intuitions carry over to more general settings, including multiplayer and repeated games. Similarly, throughout most of the main text, we study a baseline model based on one particular learning rule, called introspection dynamics (10–17). This dynamics is particularly convenient to work with because it allows for a fully analytical treatment. However, further below

(and in more detail in *SI Appendix*), we discuss to which extent the respective results generalize to other learning rules.

Once a player's sensitivity is an evolvable trait, one might expect an evolutionary dynamics toward ever-increasing sensitivities. After all, more sensitive learners are better at identifying strategies with high payoffs. Hence, such learners might have a fitness advantage. Remarkably, however, we do not find such a runaway dynamics for all games. Instead, in many snowdrift games, the players' sensitivities converge to a finite value. Moreover, in some stag-hunt games, we observe evolutionary branching. Here, an initially monomorphic population splits into two or more subpopulations, each with a different sensitivity. Interestingly, we obtain these results without assuming that more sensitive learners exert more cognitive effort and hence pay some complexity cost. Instead, we obtain these results because a noisy learning process can turn out to be a strategic advantage. In this way, our study contributes to a growing literature exploring how strategic considerations can explain behaviors that are seemingly at odds with sensible decision-making (37–41).

## Model and Results

Our model assumes a separation of time scales. In the short run, players have fixed but possibly different payoff sensitivities. They repeatedly engage in a given one-shot game, and they learn

to adopt more profitable strategies over time. We model this short-run learning process with introspection dynamics (10–17). In the long run, the players' payoff sensitivities are allowed to evolve. We describe this long-run process with adaptive dynamics (18–21), a classical tool to study the evolution of continuous traits. In the following, we describe the two processes in turn.

**A Description of the Short-Run Dynamics.** To explore learning among players with different payoff sensitivities, we consider simple one-shot,  $2 \times 2$  games. That is, we consider two players who can choose among two possible strategies, **C** and **D** (we think of the two strategies as cooperation and defection, but this interpretation is irrelevant for our results). The choices of the two players determine their payoffs. Payoffs are assumed to be symmetric, and given by the (normalized) payoff matrix

$$\begin{array}{c|cc} & \mathbf{C} & \mathbf{D} \\ \hline \mathbf{C} & 1 & -B \\ \mathbf{D} & 1-A & 0 \end{array} \quad [1]$$

We refer to this matrix as the stage game. As usual, player 1 chooses a row of this matrix, player 2 chooses a column, and the entries represent the payoffs of player 1. Depending on the signs of  $A$  and  $B$ , we recover four major game classes: prisoner's dilemma ( $A < 0, B > 0$ ), stag-hunt game ( $A > 0, B > 0$ ), snowdrift game ( $A < 0, B < 0$ ), and harmony game ( $A > 0, B < 0$ ).

To describe how players adopt more profitable strategies over time, we use introspection dynamics (10, 11). This process assumes that the two individuals interact in the same stage game over many time steps. Each time step, a randomly chosen player is given a chance to revise its strategy. That player then compares its current payoff  $\pi$  with the payoff  $\pi'$  the player could have obtained by choosing the other strategy (keeping the coplayer's strategy fixed). As illustrated in Fig. 1A, the player switches to the other strategy with a probability that depends on the payoff difference (42–44),

$$\phi = \frac{1}{1 + e^{-\beta_i(\pi' - \pi)}}. \quad [2]$$

The parameter  $\beta_i \geq 0$  represents player  $i$ 's payoff sensitivity. For small  $\beta_i$ , this function approaches  $1/2$  irrespective of the payoff difference (Fig. 1B, Left). In this limit, players decide whether to switch strategies based on a coin toss. For large  $\beta_i$ , the function approaches a step function (Fig. 1B, Right). Here, the player always switches if the alternative strategy yields the better payoff. We iterate this basic updating process for many time steps. As a result, players repeatedly get a chance to switch their strategies (Fig. 1C, Bottom). Based on this process, we record how frequently we observe each of the four possible game outcomes over time. In the Right panel of Fig. 1C, we illustrate these overall frequencies using a black-and-white gradient. A darker shading indicates game outcomes that are observed more often.

Introspection dynamics is particularly convenient because the expected frequencies of the four game outcomes can be computed analytically for any values of  $\beta$ —contrary to other stochastic dynamics such as pairwise imitation (44) or experienced-weighted attraction learning models (32). To compute the expected frequencies of each game outcome, we represent the learning dynamics as a Markov chain (see SI Appendix for details). The Markov chain has four states, corresponding to the four possible outcomes in each time step, **CC**, **CD**, **DC**, **DD** (the first letter refers to the first player's strategy, and the second letter to the

second's). For finite  $\beta_1, \beta_2$ , this Markov chain has a unique stationary distribution  $\mathbf{u}(A, B, \beta_1, \beta_2) := (u_{CC}, u_{CD}, u_{DC}, u_{DD})$ . This distribution captures how often we observe each game outcome on average, over the course of many rounds. Based on the stationary distribution, it is straightforward to compute player 1's average payoff. To this end, we multiply the four frequencies  $u_{ij}$  with the respective stage game payoffs,

$$\Pi_{\beta_2}(\beta_1) = 1 \cdot u_{CC} - B \cdot u_{CD} + (1-A) \cdot u_{DC} + 0 \cdot u_{DD}. \quad [3]$$

The respective results become particularly simple when both players have the same payoff sensitivity,  $\beta_1 = \beta_2$ . In that case, one can show that the stationary distribution simplifies to

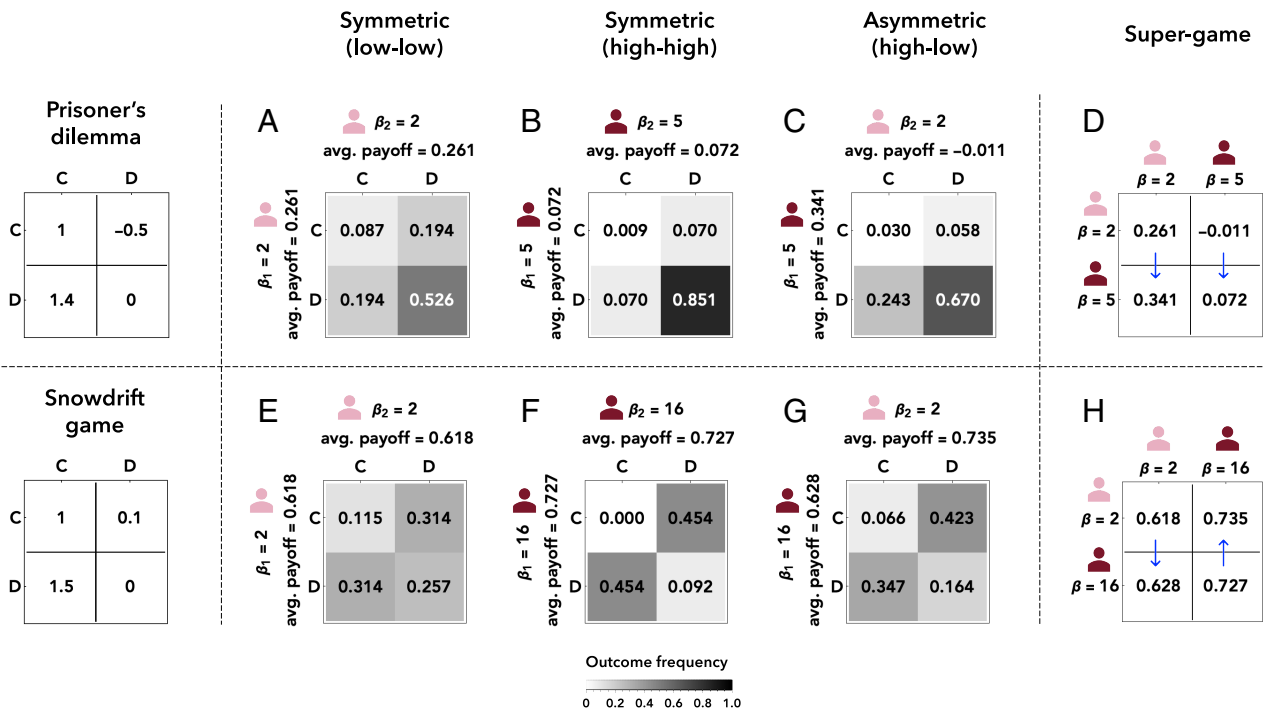
$$\mathbf{u}(A, B, \beta, \beta) = \frac{1}{2 + e^{A\beta} + e^{B\beta}} (e^{A\beta}, 1, 1, e^{B\beta}). \quad [4]$$

For weak selection ( $\beta \rightarrow 0$ ), this formula implies that all game outcomes are observed equally often,  $\mathbf{u} = (\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ . For strong selection ( $\beta \rightarrow \infty$ ), the stationary distribution depends on the game parameters  $A$  and  $B$ . For example, for the prisoner's dilemma, Eq. 4 implies that the two players defect in almost all rounds ( $u_{DD} = 1$ ); in the snowdrift game, they anticorrelate ( $u_{CD} = u_{DC} = \frac{1}{2}$ ); and in the stag-hunt game they choose the strategy that is risk-dominant ( $u_{DD} = 1$  if  $A < B$ ,  $u_{CC} = 1$  if  $A > B$ , see ref. 10).

Using the same formalism, we can also analyze players with different payoff sensitivities. In SI Appendix, we derive a general expression of the stationary distribution  $\mathbf{u}(A, B, \beta_1, \beta_2)$  as a function of  $\beta_1$  and  $\beta_2$ , which is now more complex. In the following, we use that expression to explore whether a more sensitive learner (one with a larger  $\beta$ ) would gain a long-term payoff advantage in various stage games.

In Fig. 2A–D, we consider a prisoner's dilemma. We assume the two players can either have a small or a comparably large payoff sensitivity,  $\beta \in \{2, 5\}$ . For a symmetric situation in which both players use the same  $\beta$ , we find that players learn to defect more often when  $\beta$  is large (Fig. 2A and B). This result is intuitive: The larger the payoff sensitivity, the better players are at identifying the individually optimal strategy of defecting. Once players are asymmetric, we find that it is the more sensitive learner who defects more often (Fig. 2C). As a result, this player ends up with a better average payoff.

In Fig. 2E–H, we consider a snowdrift game. For illustration, the possible sensitivities are now  $\beta \in \{2, 16\}$ . In the symmetric case among equally sensitive players, a large  $\beta$  again makes it more likely that players implement a Nash equilibrium (here, **CD** or **DC**). In the depicted example, they coordinate on a Nash equilibrium in 63% of all rounds when both have  $\beta = 2$ , but in 91% of all rounds when both have  $\beta = 16$ . Remarkably, however, once players differ in their sensitivities, it is the less sensitive (and hence more erratic) learner who obtains the larger payoff (Fig. 2G). The intuition is the following. When in their less preferred equilibrium **DC**, the less sensitive player is still relatively likely to deviate. For the given payoff values, the respective switching probability according to Eq. 2 is  $1/(1 + e^{2 \times 0.1}) = 0.45$ . This probability is smaller than  $1/2$  (because the switch reduces the player's payoff from 0.1 to 0), but it remains substantial. After this switch toward **DD**, the more sensitive player is tempted to react by playing **C**. As a result, the two players transitioned from one Nash equilibrium (**DC**, which favors the more sensitive player) to another (**CD**, which favors the less sensitive player). Interestingly, the converse transition from **CD** to **DC** is less likely because, starting from **CD**, the



**Fig. 2.** Introspection dynamics among players with different payoff sensitivities. To illustrate the impact of payoff sensitivity on the learning dynamics, we consider two stage games, a prisoner's dilemma (A–D) and a snowdrift game (E–H). In each case, we assume payoff sensitivities may take two possible values. They can either be comparably low (bright avatar) or high (dark avatar). For all possible combinations of payoff sensitivities, we depict how often the respective players would obtain one of the four possible payoffs of the respective stage game. Based on this stationary distribution, we compute the players' expected payoffs. For the prisoner's dilemma, we observe that the higher payoff sensitivity dominates the lower payoff sensitivity (indicated by blue arrows in panel D). In contrast, for the snowdrift game, no payoff sensitivity value is dominant. Instead, each player prefers to have the opposite payoff sensitivity value of their opponent (indicated by blue arrows in panel H).

less sensitive player now has more to lose by deviating (from 1.5 to 1.0).

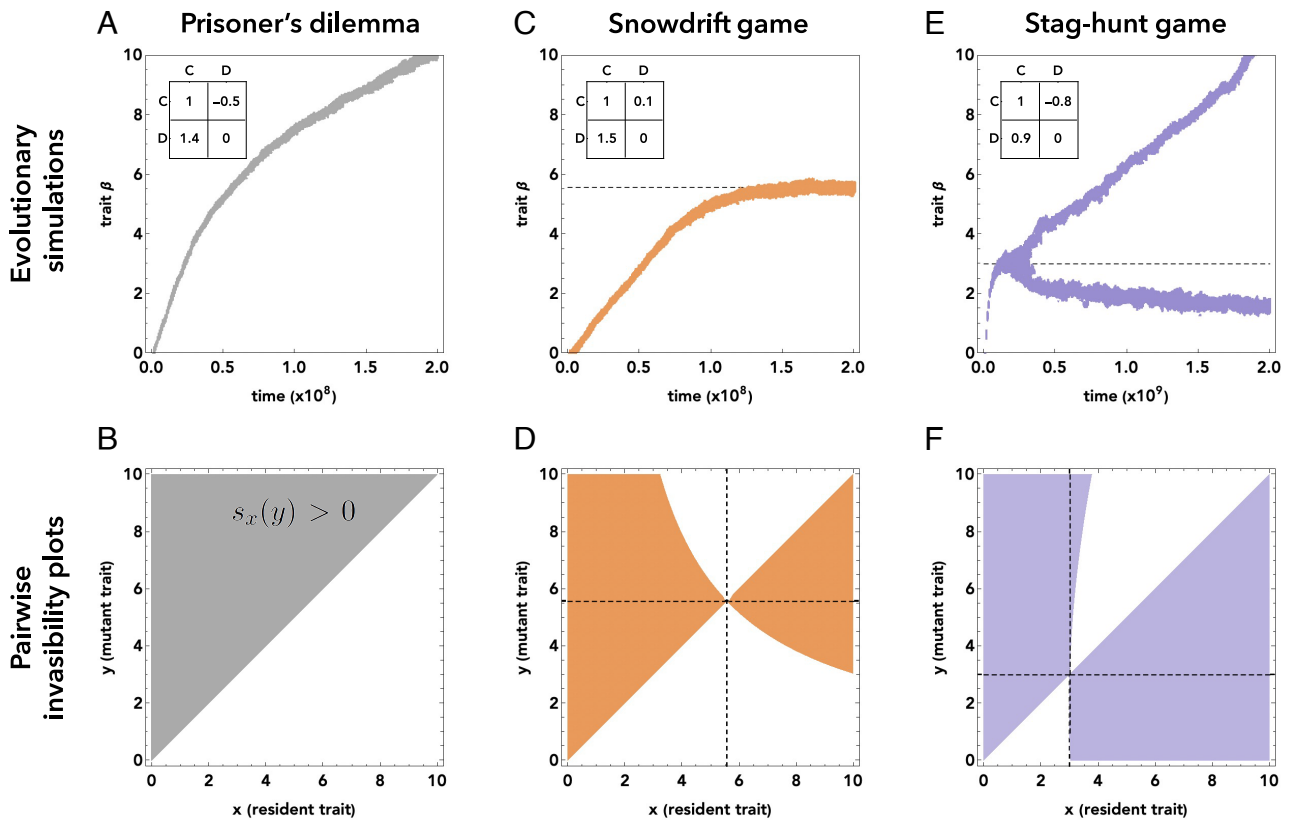
Based on the above results, we can construct a “supergame” matrix (Fig. 2 D and H). This matrix represents the interaction of two learners who may have different payoff sensitivities. The entries of the matrix depict the learners' average payoffs over the course of the learning process (as derived in the previous panels). For the prisoner's dilemma, the supergame indicates that each learner always prefers to have a higher payoff sensitivity (indicated by blue arrows). In the snowdrift game example, the supergame takes the form of an anticommodation game: Players prefer their payoff sensitivity to be high when their opponent's is low, and vice versa.

Overall, these two examples provide a proof of principle. They illustrate the nontrivial effects of asymmetric payoff sensitivities. In general, the structure of the supergame depends on both the chosen  $\beta$  values and the exact game parameters A and B. In particular, for the snowdrift game depicted in Fig. 2, one can also construct supergames in which the lower sensitivity dominates the larger one (or the other way round). Thus, these observations call for a more systematic analysis, considering more games and a larger set of possible sensitivities. We provide such an analysis in *SI Appendix*, Fig. S1. In addition, the above results raise the question of whether the players' sensitivities would necessarily move toward ever-increasing values if sensitivity were an evolving trait. We explore that issue in the following.

**A Description of the Long-Run Dynamics.** To study how the players' payoff sensitivities  $\beta$  might change over time, we use adaptive dynamics (18–20). That is, we consider a large and well-mixed population of players (Fig. 1D). Each player randomly

meets other population members to interact in a series of pairwise games. As a result, they obtain a payoff as described in the previous section on introspection dynamics. We define a player's fitness to be its average payoff across all its interaction partners. Players with larger fitness are more likely to reproduce. The offspring of a reproducing player inherits the parent's trait  $\beta$  with probability  $1 - \mu$ . With the complementary probability  $\mu$ , the offspring's payoff sensitivity is slightly mutated by adding some small noise to the parent trait. Overall, these assumptions define a dynamical process on the population level, with  $\beta$  being the evolving trait. This process is straightforward to simulate. We give a detailed description of the algorithm in *Materials and Methods*, and provide the respective code online.

Initially, we assume that the population is monomorphic, with a payoff sensitivity of  $\beta = 0$ . To gain some intuition of the further dynamics, we revisit the previous two examples. Fig. 3A shows simulations for the prisoner's dilemma. Here, we observe that  $\beta$  evolves toward ever-increasing values. This outcome is consistent with the naive expectation that less erratic learners should have a fitness advantage, and consistent with our analysis in Fig. 2 A–D. For the other example, the snowdrift game, simulations are shown in Fig. 3C. Also here, larger payoff sensitivities are initially favored. After all, positive  $\beta$  values ensure that less time is spent in off-equilibrium states (in fact, we prove in *SI Appendix*, section 1B.4 that for nondegenerate payoff matrices,  $\beta$  always increases initially). On the other hand, the supergame depicted in Fig. 2H suggests that at some point, further increases in  $\beta$  may no longer be favored, because it leads more sensitive players to spend more time in their less preferred equilibrium. Indeed, we observe the evolution of payoff sensitivities to stop at some value well below  $\beta = 6$ . That is, in contrast to our naive expectation—



**Fig. 3.** Adaptive dynamics of payoff sensitivity for three different social dilemmas. We explore the adaptive dynamics for three different social dilemmas: the prisoner's dilemma, the snowdrift game, and the stag-hunt game. For each case, we depict a realization of an individual-based simulation (*Upper panels*) and a pairwise invasibility plot (*Lower panels*). Each simulation is initialized at a monomorphic population with a payoff sensitivity of  $\beta = 0$ . Over time, mutations introduce variation in the players' payoff sensitivity. Individuals obtain payoffs by randomly interacting with other population members and learning strategies through introspection dynamics. They reproduce according to their payoff. For details, see *Materials and Methods*. Pairwise invasibility plots illustrate the dynamics of monomorphic resident populations. They display (in color) which rare mutant traits ( $y$ -axis) have a positive invasion fitness  $s_x(y)$ , given the current resident ( $x$ -axis). Colored regions above the diagonal indicate that mutants with a higher trait value than the resident are favored to invade. Colored regions below the diagonal indicate evolution toward smaller trait values. Dashed lines indicate the position of singular trait values. (*A and B*) For the depicted prisoner's dilemma, we observe a dynamics of ever-increasing payoff sensitivities. (*C and D*) In the snowdrift game, the evolving  $\beta$  values converge to a finite value. (*E and F*) In the stag-hunt game, we observe evolutionary branching. After this occurs, there are two subpopulations. Members of the first subpopulation have low  $\beta$  values, and hence they choose strategies mostly at random. Members of the second subpopulation exhibit high  $\beta$  values; they tend to best respond to their respective opponent.

but consistent with our analysis in Fig. 2 E–H—we observe the evolution of finite  $\beta$  values (or, of noisy learning).

We can gain some analytical understanding of these results by considering the respective pairwise invasibility plots (PIPs, Fig. 3 B and D). These plots depict the direction of evolution as the population size becomes infinitely large, and as mutations become sufficiently rare. In that case, we can assume the population to be monomorphic, such that all players have the same (resident) trait of  $\beta = x$ . To explore whether a mutant with a different trait  $y \neq x$  might invade, we compute the mutant's invasion fitness,

$$s_x(y) := \Pi_x(y) - \Pi_x(x). \quad [5]$$

In this formula,  $\Pi_x(y)$  is the payoff of a mutant with payoff sensitivity  $\beta = y$  in a resident population with  $\beta = x$ . This payoff is given by Eq. 3. The mutant can selectively invade the resident if and only if this invasion fitness is positive. In pairwise invasibility plots, the respective region for which  $s_x(y) > 0$  is highlighted in color. For the prisoner's dilemma, this colored area comprises the entire space above the main diagonal (Fig. 3B). This means that independent of the exact payoff sensitivity  $x$  of the resident, a mutant with trait  $y$  can invade if and only if  $y > x$ . This implies the evolution of ever-increasing payoff sensitivities.

In contrast, in the snowdrift game in Fig. 3D, mutants with a higher payoff sensitivity are only favored when the resident's sensitivity is sufficiently small (if  $x < x^*$  for some threshold  $x^*$ ). We can numerically compute this threshold by setting the so-called selection gradient equal to zero (Materials and Methods). For the depicted example, we obtain  $x^* = 5.58$ , in line with the evolutionary simulations. The pairwise invasibility plot also indicates that this singular point is evolutionarily stable. Once the population settles at a payoff sensitivity of  $\beta = x^*$ , neither a more noisy nor a less noisy learner can invade (as indicated by the white neighborhood around the vertical dashed line).

Interestingly, these two examples of an ever-increasing payoff sensitivity and an evolutionarily stable finite sensitivity are not the only possibilities. Instead, Fig. 3E shows a third scenario, using a stag-hunt game. Here, we observe evolutionary branching. At first, the monomorphic population again approaches a singular point at  $x^* = 3.01$ . Once there, both mutant types can invade, those that are more and less noisy learners (indicated by the colored neighborhood of the vertical dashed line in Fig. 3F). This is the fingerprint of a branching point. From here, the population starts diverging in opposite directions. There is a subpopulation for which sensitivities increase, and another one for which they decrease. The subsequent dynamics is more complex to describe

analytically; we do so in *SI Appendix*, section 1B.3 and Fig. S2. In the main text, we provide some intuition for why branching occurs in the first place. To this end, it is again instructive to look at the stationary distribution of introspection dynamics, for two different sensitivities  $\beta$  (*SI Appendix*, Fig. S3 A–D). When both players use the same  $\beta$ , a larger  $\beta$  makes players choose the less efficient **DD** equilibrium more often (because defection is risk-dominant). If one player instead has a low  $\beta$ , both players occasionally coordinate on the more efficient **CC** equilibrium, while often avoiding harmful miscoordination on the **CD**-outcome (*SI Appendix*, section 2C).

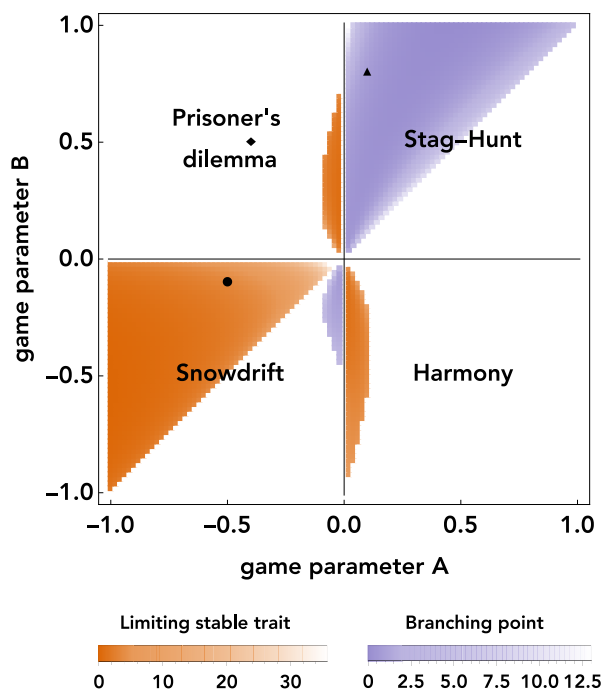
Once the players' sensitivity is allowed to evolve, its dynamics also has important implications on the population's average payoff. For example in the prisoner's dilemma, as individuals become more sensitive learners, they increasingly adopt the individually optimal strategy of defection. As a result, overall payoffs decrease in time (*SI Appendix*, Fig. S4, Left column). For the other two examples, the long-run effects on payoffs are more advantageous. Both for the snowdrift and the stag-hunt game, we observe an initially negative but eventually positive trend in the players' payoffs (*SI Appendix*, Fig. S4, Middle and Right column). In particular, the evolutionary branching in the stag-hunt game turns out to be both individually and collectively beneficial.

So far, we have studied the evolution of payoff sensitivity for a few isolated  $2 \times 2$  games. To identify more general patterns, in a next step, we systematically vary the parameters  $A$  and  $B$  of the stage game. For each resulting game, we compute the numerically exact position of the smallest singular point (if such a point exists) and its stability properties. The result is Fig. 4. Here, white regions reflect cases without a singular point. For the corresponding stage games, we predict sensitivities to increase indefinitely. Colored regions correspond to cases with at least one finite singular point. Here, we distinguish orange regions (where the smallest point is evolutionarily stable) and purple regions (where we predict branching). Based on this figure, we can make a number of observations.

First, even though most prisoner's dilemmas lead to ever-increasing payoff sensitivities (as in Fig. 3A), there are exceptions. These exceptions occur if the absolute value  $|A|$  is small (if there is only a small temptation to defect when the coplayer cooperates). These games permit a finite evolutionarily stable trait value of  $\beta$ . The underlying mechanism is different from the mechanism in the snowdrift game discussed earlier (Fig. 2 E–H). Because  $|A|$  is small compared to  $B$ , a coplayer is more likely to occasionally switch to cooperation if the focal player cooperates. This positive influence on the coplayer's cooperation rate can make it worthwhile to have a finite  $\beta$ , and to cooperate occasionally (*SI Appendix*, Fig. S3 E–H).

Second, in the snowdrift game, we observe evolution toward a finite payoff sensitivity for all games with  $B > A$ . In addition, as we consider a wider range of parameter values than considered in Fig. 4, we find another region exhibiting evolution toward finite  $\beta$ , see *SI Appendix*, Fig. S6. This second region is given by  $B < A - 1$ ; we provide a derivation of these conditions in *SI Appendix*. In both regions, the mechanism favoring finite payoff sensitivities is the same: By reducing their sensitivity, learners ensure they spend more time in their individually preferred equilibrium.

Third, in the stag-hunt game, we only observe evolutionary branching when it is risk-dominant to defect (that is, for  $A < B$ ). If instead cooperation is risk-dominant, the prevalence of **CC** increases as a function of the players' payoff sensitivity. Mutual cooperation is the optimal outcome for both players; hence, an



**Fig. 4.** Evolutionary dynamics across all social dilemmas. So far, we described three special instances of games; here we systematically repeat this analysis for all  $2 \times 2$  games. For each pair of game parameters  $A$  and  $B$ , we numerically check whether there exists a finite singular trait value  $\beta^*$ . If it exists, we explore whether the respective trait value is evolutionarily stable (orange) or a branching point (purple). The color gradient represents the position of the singular trait: the lighter, the larger the value of  $\beta^*$ . Regions without a finite singular trait are left white. Black symbols indicate the position of the three examples displayed in Fig. 3. Note that here we report the lowest singular point (the one which is reached when the process starts at  $\beta = 0$ ). However, there are games that permit two singular points. We show those in *SI Appendix*, section 2E.

ever-increasing payoff sensitivity is individually and mutually beneficial.

Finally, in the harmony game quadrant, we observe a small region that permits the evolution of finite payoff sensitivities. The existence of this region is particularly puzzling. After all, in harmony games, cooperation is both individually and collectively optimal. We discuss one of these counterintuitive cases in *SI Appendix*, Fig. S3 I–L. For this case, the supergame for two given sensitivities can take the form of a prisoner's dilemma: Collectively, the two players reach better decisions when they both choose the larger  $\beta$  value; yet individually, a player may deviate toward a smaller payoff sensitivity (as this incentivizes the more sensitive player to cooperate even more often). Interestingly, however, for each of these harmony games with a finite singular point, there exists in fact a second singular point. Hence, the adaptive dynamics for these harmony games is bistable. If players start out with small payoff sensitivities, they converge to a finite ESS. If their initial sensitivities are sufficiently large, however, their sensitivities further increase indefinitely (*SI Appendix*, Fig. S5)—perhaps the more intuitive outcome in harmony games.

**Beyond Pairwise Games.** Above, we documented several cases with an unexpected dynamics (e.g., evolution toward finite payoff sensitivities or evolutionary branching). In each case, the dynamics occurred because at least one player gained some strategic advantage from learning more erratically. In general, however, one might expect that these strategic advantages are

strongest in pairwise games. Among two players, one player's erratic behavior tends to have a comparably strong effect on how the coplayer might react. This effect could wane in groups of size  $N > 2$ . To further explore the robustness of our results, in *SI Appendix*, we study the dynamics of three classes of public goods games. The three classes differ in the shape of the public good's benefit function. Benefits are either linear, they show diminishing returns (45), or they take the form of a threshold function (46, 47). While the computational steps involved in analyzing the adaptive dynamics are the same as before, the mathematical expressions now become more intricate. Importantly, however, the qualitative dynamics are similar to before. Depending on the payoff function, again we recover the three previously observed cases of indefinite evolution, convergence to a finite evolutionarily stable state, or evolutionary branching (*SI Appendix*, Fig. S7, for  $N = 3$ ).

**Beyond Introspection Dynamics.** In all our previous examples, we studied the evolution of finite payoff sensitivities under the assumption that individuals learn their strategies based on introspection dynamics. However, the learning literature knows of many other learning rules. To explore how our results generalize, in *SI Appendix*, section 3, we numerically explore the adaptive dynamics of four other rules. All four rules can be regarded as special cases of experience-weighted attraction learning (32–34): stochastic fictitious play, payoff-informed learning, and two different variants of reinforcement learning. In each case, we explore the evolution of the player's motivation sensitivity  $\lambda$  (which plays an analogous role to the parameter  $\beta$  in introspection dynamics).

In general, for a given stage game, the four learning rules give rise to pairwise invasibility plots that can be rather different from each other, and different from introspection dynamics (*SI Appendix*, Figs. S12–S15). However, for each learning rule we find stage games for which  $\lambda$  converges toward a finite value. Again, these cases are particularly common for stage games with more than one equilibrium (the snowdrift game and the stag-hunt game). In those cases, a finite  $\lambda$  can again evolve as a means to settle at the player's more preferred equilibrium (*SI Appendix*, Figs. S9–S11). In our view, the mere existence of these finite singular points (across all considered learning rules) is remarkable. It suggests a surprisingly pervasive strategic advantage of noisy learning in certain types of stage games.

## Discussion

Evolutionary game theory explores how strategies and traits change over time (4–6). Traditionally, such an evolutionary approach can be motivated in two ways. On the one hand, the respective dynamical equations may describe long-run biological evolution. Here, individuals with better strategies are assumed to reproduce more often (48, 49). On the other hand, evolutionary models can also be interpreted as a representation of an individual's short-run learning process. Such models assume that better strategies are more readily adopted (42–44). Herein, we blend these two interpretations. We ask how a long-term biological process may shape the effectiveness of individual learning. To this end, we study the evolution of a key quantity in stochastic learning models, the players' payoff sensitivity (which is often referred to as "selection strength" in the related literature). This quantity determines to which extent payoff advantages sway individuals to change their strategies during the learning process. When individuals are comparably insensitive, they tend to ignore

their strategies' payoffs (22–26). Here, strategy choices are mostly determined by chance. As payoff sensitivities become larger, individuals become increasingly swayed by (even minor) payoff differences.

Motivated by these observations, herein we interpret payoff sensitivity as an individual trait that is related to the amount of noise in an individual's updating process. We use this perspective to study the evolution of noisy learning. To this end, we consider a model with two time scales. In the short run, individuals have fixed but possibly different payoff sensitivities. We explore the resulting learning processes with introspection dynamics (10–17); for the  $2 \times 2$  games we focus on, this process is equivalent to the logit-response dynamics studied in economics (11, 50). We use this short-term perspective to ask whether individuals who learn less erratically are able to take advantage of their opponents.

In the long run, we also allow payoff sensitivities to evolve. Here, naively one might expect ever-increasing payoff sensitivities. After all, large payoff sensitivities allow individuals to better distinguish profitable strategies from unprofitable ones. Surprisingly, however, we identify several important game classes in which payoff sensitivity converges toward a finite value. When learning more erratically, individuals occasionally use strategies that run counter to their immediate interests. In turn, players who are more sensitive to payoffs are swayed to react and to coordinate on an equilibrium that eventually benefits the noisier learner. These findings have interesting parallels in the biology literature on the red-king effect (51–53). That effect suggests that in mutualisms among two coevolving species, it can be the slower-evolving species that eventually has an evolutionary advantage.

Our model is also directly related to two interesting studies by Szabó, Vukov, and Szolnoki (35, 36). They also consider a model in which two traits coevolve. These traits are the individuals' strategies and a noise parameter that is inversely related to our payoff sensitivity. Our work deviates from these studies in many ways. For example, we look at well-mixed populations while they study behavior on a regular lattice; we treat payoff sensitivity as a continuous variable while they consider a discrete set of possible noise parameters; we explore the entire range of  $2 \times 2$  games, whereas they consider a one-dimensional subspace of games. Most importantly, however, they consider simulations in which strategies and the noise parameter evolve at the same rate. Instead, we assume a separation of time scales. As a result, they do not address most of the questions that are central to our approach. For example, because they do not allow for lasting differences in an individual's learning capabilities, they do not explore to which extent more sensitive learners are able to outperform their opponents. This question plays a crucial role in our setup, and it drives all of our subsequent results (such as the emergence of evolutionary branching, which they do not observe).

We have already sketched possible generalizations of our baseline model. In particular, in *SI Appendix* we provide more details on how the model can be extended to cover multiplayer interactions (*SI Appendix*, section 2G) or alternative learning rules (*SI Appendix*, section 3). Several other extensions are possible too. For example, herein, we considered individuals who engage in a series of one-shot (nonrepeated) games. In *SI Appendix*, section 2H, we sketch instead how to capture strategies of direct reciprocity (54) within our setup. To this end, we consider the pairwise dynamics when individuals choose between ALLD and Tit-for-Tat. Surprisingly, we find that depending on how individuals discount the future, already these two strategies can generate all of our previously described qualitative scenarios

(SI Appendix, Fig. S8). Future work could provide a more comprehensive analysis that also includes other repeated-game strategies. Other more complex setups that one could study within our framework are asymmetric games (12, 14, 15), and heterogeneous environments (55, 56). Such an approach could help us understand how people have developed psychological mechanisms to cope with the variability of situations they face.

Yet another interesting direction would be to study the evolution of noisy learning in structured rather than well-mixed populations. A framework of adaptive dynamics in such populations is readily available (57, 58). The literature shows that network structure strongly affects the way individuals cooperate (59–61). With a framework similar to ours, one could explore how network structure affects the way individuals learn. This work could investigate, for example, whether heterogeneous networks promote heterogeneity in people’s learning mechanisms. Finally, our results can also find future applications in multi-agent reinforcement learning. Specifically, one could conceive extensions where agents dynamically adapt their learning rates as a function of the dilemmas faced (62).

## Materials and Methods

We provide a detailed account of both the short-run learning dynamics and the long-run evolutionary dynamics in SI Appendix. In the following, we briefly sketch our methods for the long-run evolutionary dynamics. For that, we use a combination of individual-based simulations and a theoretical analysis in the infinite population limit.

**Evolutionary Simulations.** Our evolutionary simulations use a similar setup as Döbeli et al. (63). We start with a monomorphic population with trait  $\beta = 0$ . At each time step, two individuals,  $i$  and  $j$ , are randomly chosen from the population. These two individuals then compete for reproduction. To model this competition, each of the two individuals independently engages in a sample of three encounters with other random individuals. For each encounter, their payoff is computed with Eq. 3. By averaging over the three realizations, we compute each player’s average payoff  $\pi_i$  and  $\pi_j$ . With probability  $1/(1+\exp[-30(\pi_i-\pi_j)])$ , individual  $i$  then reproduces and  $j$  dies. Otherwise, individual  $j$  reproduces and  $i$  dies. During reproduction, there is a small chance that an offspring inherits an imperfect copy of the parent trait (with a mutation rate of  $\mu = 0.005$ ). In that case, the mutation is uniformly distributed around the parent trait in an interval of size 0.1. We iterate this simulation step a large number of times. At regular time intervals, we record the composition

of the population and the average cooperation rates and payoffs. For the simulations in Fig. 3, we consider comparably large populations ( $10^4$  in the prisoner’s dilemma and the snowdrift game;  $2 \cdot 10^4$  in the stag-hunt game).

**Theoretical Analysis.** For our analytical results, we use the adaptive dynamics formalism (18–20). Here, we assume infinitely large populations and rare mutations of infinitesimal magnitude. Under these assumptions, we can focus on mutant traits close to the resident, such that  $y \approx x$ . A key quantity to study such mutants is the *local fitness gradient* or *selection gradient*  $D(x)$ , defined by

$$D(x) := \left[ \frac{\partial s_x(y)}{\partial y} \right]_{y=x}. \quad [6]$$

The selection gradient determines the direction of evolutionary change. When  $D(x)$  is positive (negative), mutants with a slightly higher (lower) trait value than  $x$  replace the resident population. In that case, the resident trait value  $x$  grows (decreases). A point  $x^*$  that satisfies  $D(x^*) = 0$  is called a *singular point*. A singular point is called *convergence stable* if

$$\left[ \frac{dD(x)}{dx} \right]_{x=x^*} < 0. \quad [7]$$

This condition determines whether the singular point can be reached at all, starting from a resident population sufficiently close by. Moreover, the singular point is *evolutionary stable* if

$$\left[ \frac{\partial^2 s_x(y)}{\partial y^2} \right]_{x=y=x^*} < 0. \quad [8]$$

In that case, the fitness function has a local maximum at  $x^*$ , and the singular point corresponds to an evolutionary endpoint. If the singular point is at a fitness minimum, however, this may indicate the presence of an *evolutionary branching point*. For Fig. 4, we verify the respective conditions numerically.

**Data, Materials, and Software Availability.** Simulations have been run with MATLAB. For our analytical results, we used Mathematica. The respective simulation data and source code can be found in this [OSF repository](#) (64).

**ACKNOWLEDGMENTS.** M.C.C. and C.H. acknowledge generous support by the European Research Council Starting Grant 850529: E-DIRECT. F.P.S. acknowledges funding from the Dutch Research Council (NWO) through the project with file number OCENW.M.22.322 of the research programme Open Competition Domain Science. Open access funding provided by the Max Planck Society.

- M. W. Macy, A. Flache, Learning dynamics in social dilemmas. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 7229–7236 (2002).
- T. Galla, J. D. Farmer, Complex dynamics in learning complicated games. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 1232–1236 (2013).
- D. Fudenberg, D. Levine, *The Theory of Learning in Games* (MIT Press, Cambridge MA, 1998).
- M. A. Nowak, K. Sigmund, Evolutionary dynamics of biological games. *Science* **303**, 793–799 (2004).
- M. Broom, J. Rychtář, *Game-Theoretical Models in Biology* (Chapman and Hall/CRC, 2013).
- A. Traulsen, N. E. Glynatsi, The future of theoretical evolutionary game theory. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **378**, 20210508 (2023).
- J. M. McNamara, C. E. Gasson, A. I. Houston, Incorporating rules for responding into evolutionary games. *Nature* **401**, 368–371 (1999).
- J. M. McNamara, Towards a richer evolutionary game theory. *J. R. Soc. Interface* **10**, 20130544 (2013).
- O. Leimar, J. M. McNamara, Learning leads to bounded rationality and the evolution of cognitive bias in public goods games. *Sci. Rep.* **9**, 1–9 (2019).
- M. C. Couto, S. Giaimo, C. Hilbe, Introspection dynamics: A simple model of counterfactual learning in asymmetric games. *New J. Phys.* **24**, 63010 (2022).
- M. C. Couto, S. Pal, Introspection dynamics in asymmetric multiplayer games. *Dyn. Games Appl.* **13**, 1256–1285 (2023).
- O. Hauser, C. Hilbe, K. Chatterjee, M. Nowak, Social dilemmas among unequals. *Nature* **572**, 524–527 (2019).
- A. McAvoy, J. Kates-Harbeck, K. Chatterjee, C. Hilbe, Evolutionary instability of selfish learning in repeated games. *PNAS Nexus* **1**, 1–15 (2022).
- X. Wang et al., Cooperation and coordination in heterogeneous populations. *Philos. Trans. R. Soc. B* **378**, 20210504 (2023).
- V. Huebner, M. Staab, C. Hilbe, K. Chatterjee, M. Kleshnina, Efficiency and resilience of cooperation in asymmetric social dilemmas. *Proc. Natl. Acad. Sci. U.S.A.* **121**, e2315558121 (2024).
- L. Schmid, C. Hilbe, K. Chatterjee, A. N. Martin, Direct reciprocity between individuals that use different strategy spaces. *PLoS Comput. Biol.* **18**, 1–29 (2022).
- M. A. Ramirez, M. Smerlak, A. Traulsen, J. Jost, Diversity enables the jump towards cooperation for the Traveler’s Dilemma. *Sci. Rep.* **13**, 1–9 (2023).
- S. A. H. Geritz, J. A. J. Metz, E. Kisdi, G. Meszéna, Dynamics of adaptation and evolutionary branching. *Phys. Rev. Lett.* **78**, 2024–2027 (1997).
- S. A. H. Geritz, E. Kisdi, G. Meszéna, J. A. J. Metz, Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evol. Ecol. Res.* **12**, 35–57 (1998).
- J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, UK, 1998).
- Å. Brännström, J. Johansson, N. von Festerberg, The Hitchhiker’s guide to adaptive dynamics. *Games* **4**, 304–328 (2013).
- M. A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).
- C. E. Tarnita, H. Ohtsuki, T. Antal, F. Fu, M. A. Nowak, Strategy selection in structured populations. *J. Theor. Biol.* **259**, 570–581 (2009).
- A. McAvoy, A. and Hauert., Structure coefficients and strategy selection in multiplayer games. *J. Math. Biol.*, 1–36 (2015).
- B. Allen et al., Evolutionary dynamics on any population structure. *Nature* **544**, 227–230 (2017).

26. Q. Su, A. McAvooy, J. B. Plotkin, Evolution of cooperation with contextualized behavior. *Sci. Adv.* **8**, 1–11 (2022).
27. C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, K. Sigmund, Via freedom to coercion: The emergence of costly punishment. *Science* **316**, 1905–1907 (2007).
28. K. Sigmund, H. De Silva, A. Traulsen, C. Hauert, Social learning promotes institutions for governing the commons. *Nature* **466**, 861–863 (2010).
29. J. García, A. Traulsen, The structure of mutations and the evolution of cooperation. *PLoS One* **7**, e35287 (2012).
30. D. Fudenberg, M. A. Nowak, C. Taylor, L. A. Imhof, Evolutionary game dynamics in finite populations with strong selection and weak mutation. *Theor. Popul. Biol. Sci.* **70**, 352–363 (2006).
31. D. Fudenberg, L. A. Imhof, Monotone imitation dynamics in large populations. *J. Econ. Theor.* **140**, 229–245 (2008).
32. C. Camerer, T. H. Ho, Experience-weighted attraction learning in normal form games. *Econometrica* **67**, 827–874 (1999).
33. S. Dridi, L. Lehmann, On learning dynamics underlying the evolution of learning rules. *Theor. Popul. Biol. Sci.* **91**, 20–36 (2014).
34. M. Pangallo, J. B. Sanders, T. Galla, J. D. Farmer, Towards a taxonomy of learning dynamics in  $2 \times 2$  games. *Games Econom. Behav.* **132**, 1–21 (2022).
35. G. Szabó, A. Szolnoki, J. Vukov, Selection of dynamical rules in spatial Prisoner's Dilemma games. *Epl* **87**, e18007 (2009).
36. A. Szolnoki, J. Vukov, G. Szabó, Selection of noise level in strategy adoption for spatial social dilemmas. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **80**, 1–10 (2009).
37. W. Güth, H. Kliemt, The indirect evolutionary approach: Bridging the gap between rationality and adaptation. *Ration. Soc.* **10**, 377–399 (1998).
38. S. Huck, J. Oechssler, The indirect evolutionary approach to explaining fair allocations. *Games Econom. Behav.* **28**, 13–24 (1996).
39. E. Akçay, J. Van Cleve, M. W. Feldman, J. Roughgarden, A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19061–19066 (2009).
40. I. Alger, Evolutionarily stable preferences. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **378**, 20210505 (2023).
41. G. Wang, Q. Su, L. Wang, J. B. Plotkin, The evolution of social behaviors and risk preferences in settings with uncertainty. *Proc. Natl. Acad. Sci. U.S.A.* **121**, 1–10 (2024).
42. L. E. Blume, The statistical mechanics of strategic interaction. *Games Econom. Behav.* **5**, 387–424 (1993).
43. G. Szabó, C. Tóke, Evolutionary Prisoner's Dilemma game on a square lattice. *Phys. Rev. E* **58**, 69–73 (1998).
44. A. Traulsen, M. A. Nowak, J. M. Pacheco, Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 011909 (2006).
45. C. Hauert, F. Michor, M. A. Nowak, M. Doebeli, Synergy and discounting of cooperation in social dilemmas. *J. Theor. Biol.* **239**, 195–202 (2006).
46. M. Archetti, I. Scheuring, Review: Evolution of cooperation in one-shot social dilemmas without assortment. *J. Theor. Biol.* **299**, 9–20 (2012).
47. J. M. Pacheco, F. C. Santos, M. O. Souza, B. Skyrms, Evolutionary dynamics of collective action in n-person stag hunt dilemmas. *Proc. R. Soc. B* **276**, 315–321 (2009).
48. J. Maynard Smith, G. R. Price, The logic of animal conflict. *Nature* **246**, 15–18 (1973).
49. J. Hofbauer, P. Schuster, K. Sigmund, Game dynamics in mendelian populations. *Biol. Cybern.* **43**, 51–57 (1982).
50. C. Alós-Ferrer, N. Netzer, The logit-response dynamics. *Games Econom. Behav.* **68**, 413–427 (2010).
51. C. T. Bergstrom, M. Lachmann, The Red King Effect: When the slowest runner wins the coevolutionary race. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 593–598 (2003).
52. J. A. Damore, J. Gore, A slowly evolving host moves first in symbiotic interactions. *Evolution* **65**, 2391–2398 (2011).
53. C. Veller, L. K. Hayward, C. Hilbe, M. A. Nowak, The Red Queen and King in finite populations. *Proc. Natl. Acad. Sci. U.S.A.* **114**, E5396–E5405 (2017).
54. J. García, M. van Veelen, No strategy can win in the repeated prisoner's dilemma: Linking game theory and computer simulations. *Front. Robot. AI* **5**, 102 (2018).
55. D. Balliet, J. M. Tybur, P. A. Van Lange, Functional interdependence theory: An evolutionary account of social situations. *Pers. Soc. Psychol. Rev.* **21**, 361–388 (2017).
56. M. Colnaghi, F. P. Santos, P. A. M. Van Lange, D. Balliet, Adaptations to infer fitness interdependence promote the evolution of cooperation. *Proc. Natl. Acad. Sci. U.S.A.* **120**, e2312242120 (2023).
57. B. Allen, M. A. Nowak, U. Dieckmann, Adaptive dynamics with interaction structure. *Am. Nat.* **181**, E139–E163 (2013).
58. C. Hauert, M. Doebeli, Spatial social dilemmas promote diversity. *Proc. Natl. Acad. Sci. U.S.A.* **118**, e2105252118 (2021).
59. H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, A simple rule for the evolution of cooperation on graphs. *Nature* **441**, 502–505 (2006).
60. F. C. Santos, M. D. Santos, J. M. Pacheco, Social diversity promotes the emergence of cooperation in public goods games. *Nature* **454**, 213–216 (2008).
61. M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L. M. Floría, Y. Moreno, Evolutionary dynamics of group interactions on structured populations: A review. *J. R. Soc. Interface* **10**, 20120997 (2013).
62. M. Bowling, M. Veloso, Multiagent learning using a variable learning rate. *Artif. Intell.* **136**, 215–250 (2002).
63. M. Doebeli, C. Hauert, T. Killingback, The evolutionary origin of cooperators and defectors. *Science* **306**, 859–862 (2004).
64. M. C. Couto, F. P. Santos, C. Hilbe, Evolution of noisy learning in games. Open Science Framework. <https://doi.org/10.17605/OSF.IO/UKCNP>. Accessed 20 March 2026.