SCIENTIFIC REP⚙RTS

OPEN

An optimal strategy to solve the Prisoner's Dilemma

Alessandro Bravetti[1] & Pablo Padilla[1,2]

Cooperation is a central mechanism for evolution. It consists of an individual paying a cost in order to benefit another individual. However, natural selection describes individuals as being selfish and in competition among themselves. Therefore explaining the origin of cooperation within the context of natural selection is a problem that has been puzzling researchers for a long time. In the paradigmatic case of the Prisoner's Dilemma (PD), several schemes for the evolution of cooperation have been proposed. Here we introduce an extension of the Replicator Equation (RE), called the Optimal Replicator Equation (ORE), motivated by the fact that evolution acts not only at the level of individuals of a population, but also among competing populations, and we show that this new model for natural selection directly leads to a simple and natural rule for the emergence of cooperation in the most basic version of the PD. Contrary to common belief, our results reveal that cooperation can emerge among selfish individuals because of selfishness itself: if the final reward for being part of a society is sufficiently appealing, players spontaneously decide to cooperate.

Cooperation is so important that it has been suggested as a fundamental principle of evolution, besides reproduction, mutation and selection[1–13]. However, the situation regarding the underlying mechanisms responsible for the emergence of cooperation among individuals who try to maximize their own fitness and who are competing among each other is not clear. To try to solve such complicated puzzle, several mechanisms have been proposed during the last decades, including kin selection[1,6], direct reciprocity[2,3,14], indirect reciprocity[15,16], network reciprocity[17,18], group selection[9,19], green beards[5,20], optional participation[21,22], punishment and reward[23,24], pre-commitments[25–27] and others[28–31]. All these situations reflect some specific important aspects of real social and biological interactions. However, none of them can really provide a solution to the most basic version of the paradigmatic example of the Prisoner's Dilemma (PD): here one imagines that two people that are suspected of having committed a joint crime are caught by the police and confined into different rooms, without the possibility to communicate. Each of them is offered the possibility to confess the crime and defect his partner in exchange for a reduced sentence. If only one defects, the other will get the full sentence. If they defect, both will have the sentence reduced. If they cooperate between themselves and do not confess, then they will immediately be freed. The situation can be exemplified in the following payoff matrix

$$
\begin{array}{cc}
 & \begin{array}{cc} C & D \end{array} \\
\begin{array}{c} C \\ D \end{array} & \begin{pmatrix} R & S \\ T & P \end{pmatrix}
\end{array},
\tag{1}
$$

where C stands for "cooperation", D for "defection", and the entries denote the payoff for the row player. Thus, if they both cooperate, they get R points each (the "reward" for cooperation). If only one cooperates, then he gets S points (the "sucker's" payoff), while the defector gets T (the "temptation" to defect). If they both defect, they get P points each (the "punishment" for defection).

As it is well-known, in the PD the payoffs satisfy $T > R > P > S$. In this situation the best option for each player is to defect, no matter what the other player does[32]. Thus both players defect and get P points each. However, this is less than the R points they would get if they had collaborated. This is the essence of the PD: mutual cooperation leads to a higher payoff than mutual defection, but it is not a "safe" strategy, exposing the player to exploitation by a defector, and therefore each player, in an attempt to maximize his own payoff, chooses to defect. Thus apparently there is no room for cooperation between such rational agents.

[1]Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas, Universidad Nacional Autónoma de México, México City, 04510, Mexico. [2]Fitzwilliam College, University of Cambridge, Storey's Way, CB3 ODG, UK. Alessandro Bravetti and Pablo Padilla contributed equally to this work. Correspondence and requests for materials should be addressed to A.B. (email: alessandro.bravetti@iimas.unam.mx)

We remark also that in this one-shot formulation of the PD, the game is not repeated (and thus direct and indirect reciprocity do not apply), there is no link between the two prisoners (neither a genetic one as in kin selection, nor a "social" one, as in the cases of network reciprocity and group selection), nor is there a special tag that the prisoners can use (thus ruling out green beards and tag-based donation) and finally, the two prisoners cannot decide whether to play or not (and hence optional participation cannot be used). Therefore our best candidate mechanisms for the emergence of a cooperative strategy between the two prisoners do not apply in this paradigmatic example and we are left with the fundamental question: is there any other reason that can lead the two individuals to cooperate in this case?

A related unsatisfactory aspect in the formulation of evolutionary biology in terms of replicator-type dynamics with a fixed fitness landscape is the lack of adaptability: since the payoff structure is set from the beginning, e.g. (1) for the PD, and it completely determines the fitness of each strategy, the model is not flexible enough in order to incorporate changes in the payoffs which might happen over time. To overcome such difficulty, dynamical models for the coevolution of the existing species and their fitness landscape have been proposed[33–36]. This subject is also fundamentally linked to the fact that evolution is similar to an optimization process: a population evolves in order to adapt as much as possible to some external conditions (which in turn may change in time). Some works from different perspectives have been recently proposed in this context[37–39]. In particular, in[38] it has been argued that living systems adapt to the environment by performing an optimal control.

The aim of this work is to present a dynamical mechanism based on optimal control theory that has two main features: on the one side it generalizes the standard replicator-type dynamics to an evolution which can automatically adapt to changes in the fitness landscape by exerting a dynamical control on the fitness itself and on the other side it provides a simple and natural rule for the emergence of cooperation in the basic formulation of the PD presented above. Remarkably, this rule seems quite reasonable for the two prisoners as well as in biological and social interactions.

Methods

To analyze the PD from a dynamical point of view, we need to switch the perspective from game theory to evolutionary dynamics. Here the different strategies correspond to different types of individuals in a population competing by natural selection and the payoffs correspond to each type's ability to reproduce, which is called the *fitness*. For simplicity, we consider a population with only two possible types (this is all we need in order to address the PD; including more types is straightforward in principle, but the calculations get more involved very quickly). The fundamental equation of evolutionary dynamics is the Replicator Equation (RE)

$$\dot{x}_a = x_a(f_a(\mathbf{x}) - \langle \mathbf{f} \rangle) \quad a = 1, 2, \tag{2}$$

where x_a is the relative abundance (frequency) of individuals of type a, $f_a(\mathbf{x})$ is the (frequency-dependent) fitness of type a, $\langle \mathbf{f} \rangle = x_1 f_1 + x_2 f_2$ is the average fitness of the population and the overdot denotes time derivative.

For the PD, there are only two possible strategies: C or D. Thus we can label the frequency of the population adopting each strategy as x_C and x_D respectively. Moreover, the fitness of each strategy is obtained from the payoff matrix (1) by combining each payoff with the probability that the opponent chooses the corresponding strategy, thus obtaining

$$f_C(\mathbf{x}) = Rx_C + Sx_D, \quad f_D(\mathbf{x}) = Tx_C + Px_D. \tag{3}$$

Without loss of generality, we assume $S = 0$. Using the fact that $x_C + x_D = 1$, after some algebra we can rewrite the RE and the fitnesses in terms of x_C only, obtaining the following evolution equation for the frequency of cooperators

$$\dot{x}_C = -x_C(1 - x_C)[(T - R)x_C + P(1 - x_C)], \tag{4}$$

From which it is easy to deduce that there are only two fixed points $x_C = 0, 1$, and that only $x_C = 0$ is stable, that is, cooperators are dominated by defectors (cf. Fig. 1).

Thus both game theory and evolutionary dynamics agree on the fact that the best strategy in the PD is to defect. Remarkably, the standard RE always favours defectors over cooperators whenever $T > R > P > S$, independently of their relative values. Notwithstanding, one would expect that in real situations there should be a difference between e.g. the case $T = 100, R = 10, P = 9, S = 0$ and the case $T = 5, R = 4, P = 1, S = 0$. Notice also that in this classical treatment of the PD, the average fitness of the population decreases over time (see Fig. 2), meaning that the optimal strategy (defection) is optimal only with respect to the local payoff of each player, but it is not optimal with respect to the global payoff of the entire population. This is again the essence of the PD: by trying to maximize only their individual fitness, the prisoners do not achieve the best available fitness. Interestingly, this is unstable whenever the population is not isolated, but it is under evolutionary pressure by some other population, meaning that whenever selection acts both at the level of single individuals in a given population and at the level of competing populations, a strategy considering only the former aspect is destined to be suppressed by one considering both features (this is the main reason why group selection achieves cooperation[9,19]).

Let us now use this observation in order to construct a mechanism that boosts the emergence of collaboration among competing individuals. Motivated by the fact that being part of a society provides in many situations a concrete evolutionary advantage for each organism, we assume that natural evolution acts by two mechanisms: on the one side it selects those individuals inside a population with higher fitness at any given time, and on the other side it selects those populations with the higher final average fitness (we can see this process as due to the existence of two time scales, a faster one, in which selection operates at the level of individuals in a population, and a slower scale, in which selection operates among populations). We require the evolution in the faster scale

	T	R	P	S	x_C^0
Model 1	5	4	1	0	0.3
Model 2	4	3	2	0	0.2
Model 3	4	3.5	0.5	0.5	0.1

Table 1. Models used for the graphics. We report the values of the payoffs and of the initial conditions used in the different models given in Figs 1 and 2.
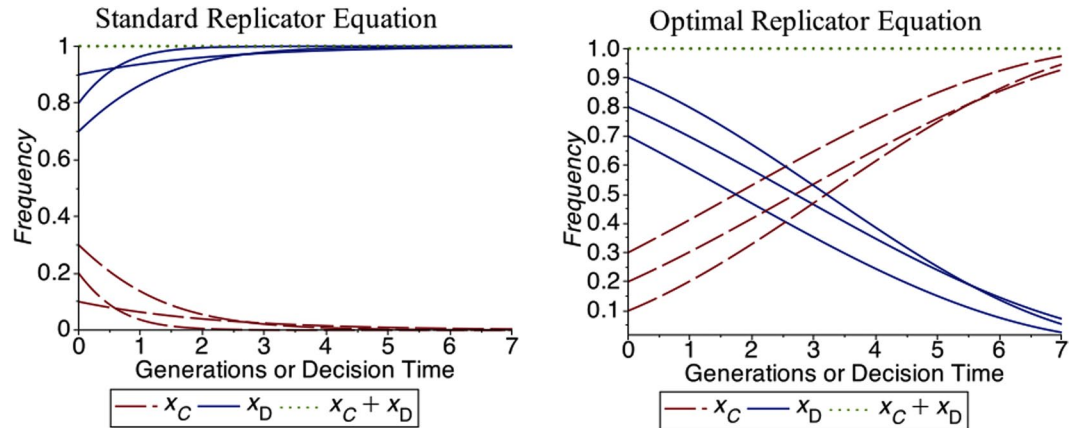

Figure 1. Evolution of cooperators (C) and defectors (D) in the PD. On the left panel we present the evolution according to the standard Replicator Equation, while the right panel is the evolution of the same system according to the Optimal Replicator Equation introduced here. The payoffs and the initial conditions used in these examples are reported in Table 1. While the standard RE predicts that cooperators disappear during evolution, the ORE shows the emergence of cooperation.
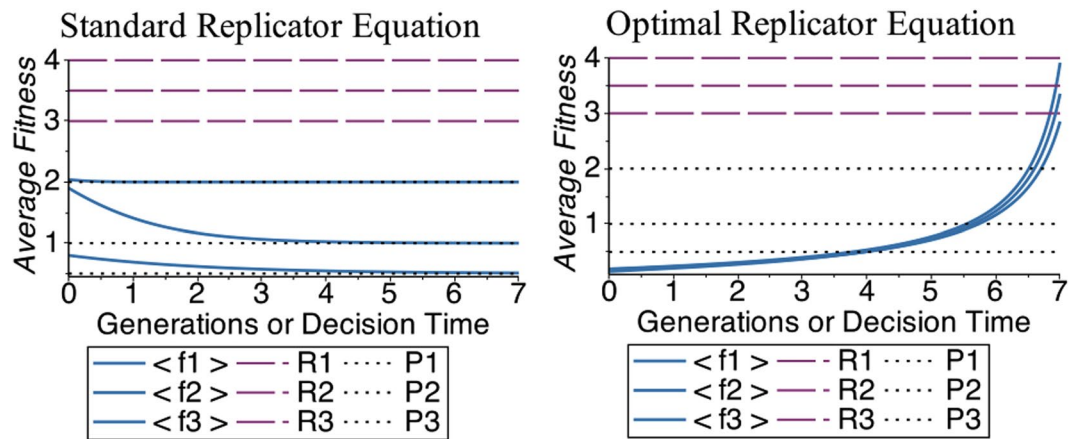

Figure 2. Evolution of the average fitness of the population in the PD. On the left panel we present the evolution according to the standard Replicator Equation, while the right panel is the evolution of the same system according to the Optimal Replicator Equation introduced here. The payoffs and the initial conditions used in these examples are reported in Table 1. While using the standard RE the average fitness of the population decreases and approaches the payoff for mutual defection (P), in the system governed by the ORE the average fitness increases and approaches the payoff for mutual cooperation (R).

to be dictated locally in time by the standard RE. Finally, in order to obtain a model sensitive to changes in the environment, we suppose that the fitness of each strategy can be regulated in response to the present state of the population (this is similar to the coevolution mechanisms proposed e.g. in refs[33,34]).

The above assumptions can be translated into the mathematical problem of maximizing a functional which consists of two terms (one corresponding to selection at the level of individuals, and one corresponding to selection at the level of populations) subject to a dynamical constraint set by the standard RE. In order to find the best strategy to realize this task, we need to solve a maximization problem (similar ideas have been put forward in[37–39]). This is the arena of Optimal Control Theory (OCT)[40–43]. By applying OCT we obtain (see Supplementary

Information) an enlarged system of equations that extends the RE (2) and describes the dynamical equations for the coevolution of the frequencies $\mathbf{x}$ in the RE and the fitnesses $\mathbf{f}$. For $a = C, D$, as in the case of the PD, this system reads

$$f_a = \frac{1}{2}p_a,$$
(5)

$$\dot{x}_a = \frac{x_a}{2}(p_a - \langle \mathbf{p} \rangle),$$
(6)

$$\dot{p}_a = \frac{p_a}{2}\left(\langle \mathbf{p} \rangle - \frac{p_a}{2} \right).$$
(7)

The additional variables p_C and p_D are usually called the *co-states* (see Supplementary Information). We call the system (5)–(7), together with the initial conditions $x_a(0) = x_a^0$ and the terminal conditions

$$p_a(\tau) = \frac{\partial g(\mathbf{x}(\tau))}{\partial x_a}, \qquad g(\mathbf{x}) := \langle \mathbf{f} \rangle$$
(8)

the *Optimal Replicator Equation* (ORE). In the following we show that the ORE leads directly to a simple rule for the emergence of cooperation in the PD.

Results

Let us consider now the PD in terms of the ORE. As usual, we use $x_C + x_D = 1$ to simplify (6),(7) by eliminating the equation for x_D. Indeed, one can use $\dot{x}_D = -\dot{x}_C$ to rewrite the two equations in (6) as a single equation for x_C, which reads

$$\dot{x}_C = x_C(1 - x_C)\frac{p_C - p_D}{2}.$$
(9)

Thus we see that we have obtained already a simple condition for the emergence of cooperation, that is, $p_C - p_D > 0$. Since by the optimal control strategy (5), p_C and p_D correspond to the fitness of cooperators and defectors respectively, the above condition seems almost trivial at first sight: cooperation emerges whenever $f_C > f_D$. Nevertheless, this condition is not trivial in this case, at least for two reasons: firstly, it is obtained as the result of an optimal strategy; secondly, and most importantly, because the two variables p_C and p_D are *dynamical*, with dynamical equations (7) and terminal conditions determined by the choice of the final average fitness according to (8). This aspect is crucial in the solution of the PD.

Let us return to the case of our two prisoners, with payoff matrix as in (1). Now we consider the payoffs in (1) as a *final* payoff given at time $t = \tau$ and use the ORE with the final fitness for each strategy given by (3). As before, we assume $S = 0$, thus the final average fitness is

$$g(\mathbf{x}(\tau)) = Rx_C(\tau)^2 + Tx_C(\tau)x_D(\tau) + Px_D(\tau)^2,$$
(10)

and the terminal conditions (8) read

$$p_C(\tau) = (2R - T)x_C(\tau) + T,$$
(11)

$$p_D(\tau) = (T - 2P)x_C(\tau) + 2P.$$
(12)

the system of equations (7) and (9), together with the initial conditions $x_C(0) = x_C^0$ and $x_D(0) = 1 - x_C^0$ and the terminal conditions (11),(12) is our dynamical model for the PD. Using (9) and (11),(12) one can prove that for

$$2P \leq T \leq 2R,$$
(13)

the right hand side of (9) at $t = \tau$ is always greater than or equal to zero, with equality only for $x_C = 0, 1$. This means that this evolution admits only two equilibria, namely $x_C = 0, 1$ and that only $x_C = 1$ is asymptotically stable. Therefore (13) is the condition for the emergence of cooperation in the PD using the ORE. Typical numerical solutions are given in Fig. 1. As we see, contrary to the standard RE, according to the ORE cooperators take over the population.

Another difference with respect to the standard RE is that the average fitness of the population *increases* with the ORE: while the standard evolution dictated by the RE predicts that the average fitness of the population decreases, approaching the payoff of mutual defection (see Fig. 2), the ORE predicts that selfish individuals cooperate in order to maximize the final average fitness of the population, because this entails an advantage for themselves.

The reason for the cooperative behavior in the ORE is simple: using the ORE we have extended the RE to a dynamics which considers the important evolutionary advantage for each individual deriving from being in a population that has a large final average fitness. Knowing that in case of collaboration they will be able to share an important final payoff, naturally induces the two prisoners to collaborate. Interestingly, collaboration only appears whenever R is "high enough", that is, whenever $2R \geq T$, and also whenever P is "low enough", that is,

whenever $2P \geq T$. In both cases the prisoners do not find advantageous to defect, and prefer to take the risk to be exposed to exploitation rather than taking a lower payoff. They consider that in such cases the risk is worth the price (cf.[44]).

In biology we can safely assume that in many situations being part of a society guarantees a much higher probability of survival, for instance because it provides better strategies for the collection of food, or for reproduction, or for defense against predators. In this biological setting, we argue that the effect of cooperation can only be assessed at the level of the average fitness of the whole population. So, even if a tendency to cooperate might be inherited, its evolutionary advantage can only be evaluated collectively a posteriori and therefore cannot be included in a local term for the RE, but rather as a final condition. A striking example is provided by bees: in bees society, usually only queens can reproduce. Therefore an evolution based only on the ability to reproduce would lead very quickly to the disappearance of any worker bee. However, workers play an important role for finding food and defending the hive. Cooperation between queens and workers means that the former guarantee reproduction for the latter, while the latter work for the former. This leads to a huge final reward, that is, the conservation of the species after each generation.

Finally, let us stress that while in the case of the PD the two prisoners can be aware of the final reward for collaboration and therefore they can (consciously) decide to cooperate under the appropriate conditions, in a biological setting we can no longer give a similar interpretation. However, as it is usual in evolutionary dynamics, we suppose that the different individuals in the population inherit the possible traits randomly and that evolution favours only those individuals with the best traits, so that the optimization process is a consequence of natural selection and not a conscious decision in such case.

Discussion

To summarize, we have proposed a modified version of the Replicator Equation (RE), the equation governing natural selection, called the Optimal Replicator Equation (ORE), which stems from the assumption that evolution is an optimization process that on the one side selects at any given time those individuals with higher fitness and on the other side favours those populations with higher average fitness. The main motivations for the introduction of such model are the facts that the standard RE cannot account for selection on the two levels of individuals and populations and that it fails to reproduce observed situations, such as the emergence of cooperation in the Prisoner's Dilemma. Interestingly, by implementing our model (which by definition takes into account the two levels of selection among individuals and populations) to the case of the PD, we have shown that the corresponding dynamics naturally favours cooperation in the case of the basic Prisoner's Dilemma under some reasonable conditions (cf. ([13])). Our results thus open the door for an investigation of evolution and social dilemmas in terms of optimization by using the reproduction coefficients – i.e. the fitness – as control parameters. In particular, it would be interesting for future work to compare the condition for the emergence of cooperation obtained here with data from various experiments on the PD and with conditions derived from similar models[44–46]. Moreover, one can enlarge the study of the optimal strategies deriving from the ORE by considering different social dilemmas beside the PD. We expect to find results on the emergence of cooperation similar to the ones presented here, thus enforcing the idea that the ORE can be a good dynamical model for explaining the emergence of cooperation in a competitive framework.

References

1. Hamilton, W. D. The genetical evolution of social behaviour. II. *Journal of theoretical biology* **7**(1), 17–52 (1964).
2. Trivers, R. L. The evolution of reciprocal altruism. *The Quarterly review of biology* **46**(1), 35–57 (1971).
3. Axelrod, R. The emergence of cooperation among egoists. *American political science review* **75**(2), 306–318 (1981).
4. Nowak, M. A. & Sigmund, K. Tit for tat in heterogenous populations. *Nature* **355**(6357), 250 (1992).
5. Riolo, R. L., Michael, D. C. & Axelrod, R. Evolution of cooperation without reciprocity. *Nature* **414**(6862), 441 (2001).
6. West, S. A., Pen, I. & Griffin, A. S. Cooperation and competition between relatives. *Science* **296**(5565), 72–75 (2002).
7. Nowak, M. A. *et al.* Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**(6983), 646 (2004).
8. Nowak, M A. & Sigmund, K. *Evolution of indirect reciprocity*. (2005).
9. Traulsen, A. & Martin, A. N. Evolution of cooperation by multilevel selection. *Proceedings of the National Academy of Sciences* **103**(29), 10952–10955 (2006).
10. Axelrod, R M. *The evolution of cooperation: revised edition*. Basic books (2006).
11. Nowak, M. A. Five rules for the evolution of cooperation. *science* **314**(5805), 1560–1563 (2006).
12. Taylor, C. & Martin, A. N. Transforming the dilemma. *Evolution* **61**(10), 2281–2292 (2007).
13. Hagel, K. *et al.* Which risk scenarios can drive the emergence of costly cooperation? *Scientific reports* **6**, 19269 (2016).
14. Milinski, M. Tit for tat in sticklebacks and the evolution of cooperation. *nature* **325**(6103), 433–435 (1987).
15. Nowak, M. A. & Sigmund, K. *Evolution of indirect reciprocity by image scoring/the dynamics of indirect reciprocity* (1998).
16. Manfred, M., Semmann, D. & Krambeck, H.-J. Reputation helps solve the 'tragedy of the commons'. *Nature* **415**(6870), 424–426 (2002).
17. Nowak, M. A. & Robert, M. M. Evolutionary games and spatial chaos. *Nature* **359**(6398), 826–829 (1992).
18. Ohtsuki, H. *et al.* A simple rule for the evolution of cooperation on graphs. *Nature* **441**(7092), 502 (2006).
19. Perc, M. *et al.* Evolutionary dynamics of group interactions on structured populations: a review. *Journal of the royal society interface* **10**(80), 20120997 (2013).
20. Jansen, V. A. A. & Baalen, V. A. Altruism through beard chromodynamics. *Nature* **440**(7084), 663 (2006).
21. Hauert, C. *et al.* Volunteering as red queen mechanism for cooperation in public goods games. *Science* **296**(5570), 1129–1132 (2002).
22. Ghang, W. & Martin, A. N. Indirect reciprocity with optional interactions. *Journal of theoretical biology* **365**, 1–11 (2015).
23. Szolnoki, A & Perc, M. Antisocial pool rewarding does not deter public cooperation. *Proc. R. Soc. B*. Vol. 282. No. 1816. The Royal Society (2015).
24. Chen, X., Szolnoki, A. & Perc, M. Competition and cooperation among different punishing strategies in the spatial public goods game. *Physical Review E* **92**(1), 012819 (2015).
25. Han, T. A. *et al.* Good agreements make good friends. *Scientific reports* **3** (2013).
26. Pereira, L. M. & Lenaerts, T. Avoiding or restricting defectors in public goods games? *Journal of the Royal Society Interface* **12**(103), 20141203 (2015).

27. Sasaki, T. *et al.* Commitment to cooperation and peer punishment: Its evolution. *Games* **6**(4), 574–587 (2015).
28. Chen, X. *et al.* First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation. *Journal of The Royal Society Interface* **12**(102), 20140935 (2015).
29. Szolnoki, A. & Chen, X. Benefits of tolerance in public goods games. *Physical Review E* **92**(4), 042813 (2015).
30. Szolnoki, A. & Chen, X. Cooperation driven by success-driven group formation. *Physical Review E* **94**(4), 042311 (2016).
31. Chen, X. & Szolnoki, A. Individual wealth-based selection supports cooperation in spatial public goods games. Scientific reports 6 (2016).
32. Nowak, M. A. *Evolutionary dynamics*. Harvard University Press (2006).
33. Nilsson, M. & Snoad, N. Error thresholds for quasispecies on dynamic fitness landscapes. *Physical Review Letters* **84**(1), 191 (2000).
34. Wilke, C. O., Ronnewinkel, C. & Martinetz, T. Dynamic fitness landscapes in molecular evolution. *Physics Reports* **349**(5), 395–446 (2001).
35. Klimek, P., Thurner, S. & Hanel, R. Evolutionary dynamics from a variational principle. *Physical Review E* **82**(1), 011901 (2010).
36. Karev, G. P. On mathematical theory of selection: continuous time population dynamics. *Journal of mathematical biology* **60**(1), 107–129 (2010).
37. Traulsen, A., Iwasa, Y. & Martin, A. N. The fastest evolutionary trajectory. *Journal of theoretical biology* **249**(3), 617–623 (2007).
38. Chakrabarti, R. *et al.* Mutagenic evidence for the optimal control of evolutionary dynamics. *Physical review letters* **100**(25), 258103 (2008).
39. Saakian, D. B., Makar, H. G. & Hu, C.-K. Punctuated equilibrium and shock waves in molecular models of biological evolution. *Physical Review E* **90**(2), 022712 (2014).
40. Geering, H P. Optimal control with engineering applications. *Berlin Heidelberg* (2007).
41. Lenhart, S. & J. T. Workman Optimal control applied to biological models. Crc Press, 2007.
42. Evans, L. C. An introduction to mathematical optimal control theory. Lecture Notes, University of California, Department of Mathematics, Berkeley (2005).
43. Fleming, W. H. and Raymond W. Rishel. Deterministic and stochastic optimal control. Vol. 1. Springer Science & Business Media, 2012.
44. Engel, C. & Zhurakhovska, L. When is the risk of cooperation worth taking? The prisoner's dilemma as a game of multiple motives. *Applied Economics Letters* **23**(16), 1157–1161 (2016).
45. Capraro, V. A model of human cooperation in social dilemmas. *PLoS One* **8**(8), e72427 (2013).
46. Capraro, V., Jillian J. J. & David G. R. Heuristics guide the implementation of social preferences in one-shot Prisoner's Dilemma experiments. *Scientific reports* 4 (2014).

Acknowledgements

Author Contributions

A.B. and P.P. contributed to all aspects of this work.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-20426-w.

Competing Interests: The authors declare that they have no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.