Journal of Theoretical Biology 433 (2017) 64–72

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtbi

Optional interactions and suspicious behaviour facilitates trustful cooperation in prisoners dilemma

Tadeas Priklopil[a,b,*], Krishnendu Chatterjee[b], Martin Nowak[c]

[a] Department of Ecology and Evolution, University of Lausanne, Lausanne, Switzerland
[b] Institute of Science and Technology IST Austria (IST Austria), Am Campus 1, A-3400 Klosterneuburg, Austria
[c] Program for Evolutionary Dynamics, Department of Mathematics, and Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

ARTICLE INFO

Article history:
Received 12 April 2017
Revised 18 August 2017
Accepted 30 August 2017
Available online 1 September 2017

Keywords:
Evolutionary game theory
Optional interactions
Evolution of cooperation
Non-social behaviour
Partial information

ABSTRACT

In evolutionary game theory interactions between individuals are often assumed obligatory. However, in many real-life situations, individuals can decide to opt out of an interaction depending on the information they have about the opponent. We consider a simple evolutionary game theoretic model to study such a scenario, where at each encounter between two individuals the type of the opponent (cooperator/defector) is known with some probability, and where each individual either accepts or opts out of the interaction. If the type of the opponent is unknown, a trustful individual accepts the interaction, whereas a suspicious individual opts out of the interaction. If either of the two individuals opt out both individuals remain without an interaction. We show that in the prisoners dilemma optional interactions along with suspicious behaviour facilitates the emergence of trustful cooperation.

1. Introduction

Evolutionary games provide a general framework to study frequency dependent selection, where the fitness (payoff) of each individual is determined by playing a game with other individuals in the population. In the standard formulation, games between individuals are assumed compulsory in the sense that individuals cannot choose whom they encounter, and are then forced to execute their strategy with the encountered individual (e.g. Weibull, 1995). In nature, however, this is usually not the case. Various models have taken this into consideration and allowed individuals to be selective either in the form of partner choice ("pre-interaction decisions", e.g. Noë and Hammerstein, 1994; Hruschka and Henrich, 2006; Fu et al., 2008) and/or partner switching ("post-interaction decisions", e.g. Hruschka and Henrich, 2006; McNamara et al., 2008; Fu et al., 2008; Fujiwara-Greve and Okuno-Fujiwara, 2009; Izquierdo et al., 2010; Wubs et al., 2016; Zheng et al., 2017), whilst some models have allowed individuals to refuse interactions altogether ("optional interactions", e.g. Miller, 1967; Vanberg and Congleton, 1992; Orbell and Dawes, 1993; Stanley et al., 1995; Batali and Kitcher, 1995; Sherratt and Roberts, 1998; Hauert et al., 2002a; Mathew and Boyd, 2009; Ghang and Nowak, 2015). In this work we assume that opponents are chosen at random and we focus on optional interactions.

An extremely simple form of optional interactions is to accept no interactions, which is the so-called loners strategy (Brandt et al., 2006; Cardinot et al., 2016; Fowler, 2005; Hauert et al., 2002a; 2002b; 2007; Mathew and Boyd, 2009). Individuals who adopt a loners strategy opt out of all interactions and receive a fixed "loners payoff". In evolutionary games with loners along with cooperators and defectors, where cooperators and defectors are assumed to accept every interaction, the evolutionary trajectories approach a cycle between the three strategies (Hauert et al., 2002a; 2002b). This is an interesting result, particularly because in such models individuals have no information about their opponents.

While no information and no interaction represents an extreme scenario, in many situations individuals can base their decision on partial information about their opponent. Classical examples where individuals in the population have at least some information about each other are as follows: (a) models of direct reciprocity: individuals have encountered their opponent in the past (Batali and Kitcher, 1995; Castro and Toro, 2008; Kurokawa, 2017; Sherratt and Roberts, 1998; Spichtig et al., 2013; Trivers, 1971); (b) models of indirect reciprocity: the opponent has build a reputation of its past actions with other individuals (Fu et al., 2008; Ghang and Nowak, 2015; Nowak and Sigmund, 2005; 1998a; 1998b; Panchanathan and Boyd, 2003); or (c) the opponent appears or behaves a certain way before an interaction takes place that indicates its intended

* Corresponding author.
E-mail address: tadeas.priklopil@unil.ch (T. Priklopil).

http://dx.doi.org/10.1016/j.jtbi.2017.08.025
0022-5193/

actions (DeSteno et al., 2012; Frank et al., 1993; Reed et al., 2012; Yamagishi et al., 1999). For example, the ability of correctly evaluating mate selection-related strategies of other individuals is common (Andersson and Simmons, 2006; Iwasa et al., 1991; Jennions and Petrie, 1997; Zahavi, 1975). In such situations, and in contrast to loners strategy of always opting out, the decision of opting out or accepting the interaction ought to depend on the available partial information.

In this work we introduce a simple evolutionary game-theoretical model where the individuals encounter each other at random (no choice of opponents), but at each encounter they are given the option to either accept or opt out of the interaction based on partial information about their opponent. If either of the two individuals opt out, both individuals remain without an interaction. In our model the type of the opponent (cooperator or defector) is known with some fixed probability. If the type of the opponent is known, then individuals take a decision (accept or opt out) that yields a greater payoff. If the type of the opponent is not known, then individuals can be either trustful or suspicious (Panchanathan and Boyd, 2003; Sigmund, 2010). A trustful individual accepts an interaction with the trust that the opponent will provide a greater payoff than opting out, and a suspicious individual opts out of an interaction suspecting that the opponent will provide a lesser payoff than what opting out yields. The strategy of an individual is thus a combination of its type (cooperator/defector) and a decision rule that dictates whether to accept or opt out of an interaction (trustful/suspicious).

We formally introduce our modeling framework in the following section, and then as an example, study the evolution of cooperation by working out the game of prisoners dilemma in detail. We succinctly summarize our key findings below.

- First, if the probability of knowing the type of the opponent is above a certain threshold, a threshold that is given in terms of payoffs, then trustful cooperation is an ESS. A similar condition was derived in Nowak and Sigmund (1998a), Nowak and Sigmund (1998b), Suzuki and Toquenaga (2005) and Ghang and Nowak (2015). Interestingly, and in contrast to the previous findings, if opting out yields an equal or greater payoff than mutual defection, then trustful cooperation is a globally convergent ESS, i.e., trustful cooperation is reached from any initial state of the population. In particular, even an (almost) entirely defective population will be eventually replaced by trustful cooperators.
- Second, we consider that the probability of knowing the type of the opponent is below the required threshold. If opting out is at least as beneficial as mutual defection, then the evolutionary dynamics approaches a rock-paper-scissors cycle of trustful cooperation, trustful defection and suspicious cooperation. However, if opting out is strictly better than mutual defection, then for a low probability of knowing the type of the opponent, trustful cooperation, trustful defection and suspicious cooperation coexist at a globally stable equilibrium. We note that suspicious defection is always (eventually) selected against and thus eradicated form the population.

To summarize, we introduce a simple mathematically tractable model that enables us to study the interplay between social (accepted interactions) and non-social (rejected interactions) behaviour. We apply our model to the game of prisoners dilemma where we show that the option of non-social behaviour of opting out of interactions, a “natural precondition” of partner formation, allows for the emergence of (social and) cooperative behaviour. Moreover, we find that non-social behaviour together with the ability to recognise the behaviour of each other leads not only to stable cooperative populations but also to trustful behaviour that accepts interactions with potentially defective players.

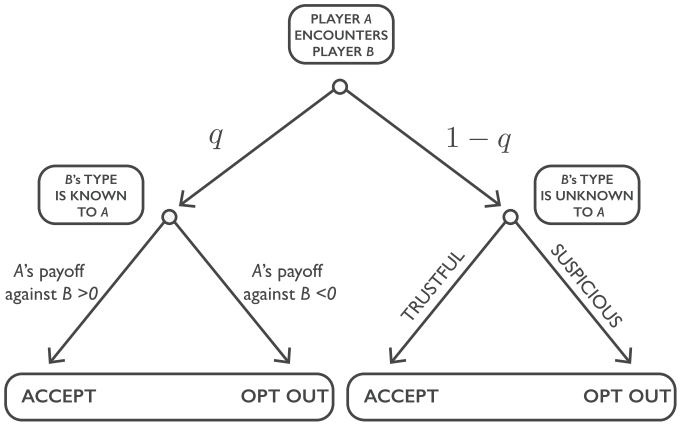

Fig. 1. Decision tree for a player of type *A*. At the top node nature decides whether the player *A* identifies the type of the encountered opponent *B* or not, which happens with probabilities q and $1-q$, respectively. If player *A* identifies the type of the encountered opponent (left branch), the player chooses the action that maximizes its payoff. Thus player *A* will accept the interaction if the payoff of *A* against *B* is greater than 0, otherwise player *A* will opt out of the interaction. If player *A* doesn't identify the type of its encountered opponent (right branch), player *A* can either be trustful or suspicious and will either accept or opt out of the interaction, respectfully.

2. Model description

Consider a large and well-mixed population with two types of players, cooperators and defectors. Players are assumed to encounter each other at random, such that at each encounter they can either accept or reject each other for an interaction. If both players accept, a game is played and a payoff is received: if both players are cooperators both receive R, if both players are defectors both receive P, and if one is a defector and the other is a cooperator then the defector receives T and the cooperator S, such that $S<P<R<T$. A game is not played if at least one of the two players rejects the interaction (opt out), in which case both players receive a payoff L, where L can be any value relative to the payoffs S, P, R, T. Without loss of generality we set $L=0$ and scale the other payoffs accordingly (SI). The payoffs S, P, R, T thus need to be reinterpreted as the difference between the particular social interaction and non-social behaviour. We note that each player knows its own type as well as the ordering of payoffs.

The decision to accept or opt out of an interaction is made based on the type of the opponent, which is known to the player with some fixed probability q. If the type of the opponent is known the decision to interact is obvious – a game that yields a greater payoff than opting out will be accepted and with a smaller payoff rejected. This is illustrated with the left branch in Fig. 1 where a player of type *A* has identified the type of the encountered opponent *B*. The question is what to do when the opponent is unknown (the right branch in Fig. 1). Since players have no information about the composition of the population (frequency distribution of cooperators and defectors) they have only two options, either *trust* that by accepting the interaction the unknown player will yield them a greater payoff than if they chose to opt out, or be *suspicious* that the interaction will be advantageous and reject the unknown opponent (Panchanathan and Boyd, 2003; Sigmund, 2010). All in all we obtain four strategies, trustful cooperation, suspicious cooperation, trustful defection and suspicious defection, keeping in mind that for some payoff configurations not all strategies are rational and hence will not be considered. For example, if mutual defection yields greater payoff than non-social behaviour $0<P$, then defectors will always receive a greater payoff by accepting an interaction, known and unknown, and thus the strategy of suspicious defection will be disregarded.

We immediately observe that the cases $0 \leq S$ and $R \leq 0$ lead to trivial evolutionary dynamics (Batali and Kitcher, 1995). If $0 \leq S$, then any interaction is at least as good as no interaction and thus all games should be accepted, and if $R \leq 0$, then cooperators receive always the maximum payoff by not interacting and so all games end up being rejected. In the first case we recover the dynamics of the prisoners dilemma with obligatory interactions where defective strategy is the evolutionary outcome. In the latter case players of both types opt out of all interactions. Thus, the task is to work out the evolutionary dynamics for the two remaining cases, $S < 0 \leq P < R < T$ and $S < P < 0 < R < T$. We remark that the non-generic case $P = 0$ is of special interest and will be considered separately, not only due to its simple evolutionary dynamics but also because a donation game, the central model in the literature of evolution of cooperation (Sigmund, 2010), falls into this category of models when the benefit of defection $T - R$ and the cost of cooperation $S - P$ are equal.

We will interpret mutual defection as some arbitrary social interaction that provides basic income P, where the potentially harmful effect of the interaction is factored in the payoff. Depending on the level of harm defection causes to the co-player, the payoff for mutual defection may be greater or smaller than the payoff for non-social behaviour. Equivalently, and this is the terminology we use throughout the paper, we say that opting out is costly when $P > 0$ and beneficial when $P < 0$.

3. Results

We will first work out a model for the two limiting cases where players have either zero information $q = 0$ or perfect information $q = 1$ about their opponents. In the following sections we will consider games with partial information $0 < q < 1$ and first deal with the special case $P = 0$ where opting out and mutual defection results in equal payoff. Lastly we solve the two remaining cases, $S < 0 < P$ where opting out is costly and $P < 0 < R$ where opting out is beneficial. For each model we analyse the evolutionary dynamics represented with a continuous-time replicator equation

$$\dot{x}_A = x_A\left(E_A - \bar{E}\right) \tag{1}$$

where the dot denotes a time derivative, x_A is the frequency and E_A is the expected payoff of strategy A, and $\bar{E} = \sum_B x_B E_B$ is the average payoff in the population.

3.1. Games with zero and perfect information

Let us first consider the case where players have zero information about the type of the opponent $q = 0$ and so all interactions are between unknown players. In both non-trivial cases $S < 0 \leq P < R < T$ and $S < P < 0 < R < T$ we have $S < 0 < R$, and so the decision for a cooperator to accept or opt out of an interaction with an (always) unknown opponent depends whether the opponent is likely to be a cooperator or a defector. If the unknown opponent is likely to be a defector it pays off to opt out, but if the opponent is likely to be a cooperator it pays off to accept the interaction. We thus need to consider both suspicious and trustful cooperators, where suspicious cooperators opt out of all interactions, while trustful cooperators accept every interaction. Similarly, if $P < 0 < R$ defectors may either be suspicious and opt out of all interactions or be trustful and always defect. However, for $S < 0 \leq P$ all defectors ought to be trustful and accept every interaction. In this case suspicious defectors will not be considered. We thus need to consider only three simple strategies, suspicious strategies (i.e. suspicious cooperators and for $P < 0 < R$ also suspicious defectors) who opt out of every interaction, trustful cooperators and trustful defectors who accept every interaction. The expected payoff for suspicious strategies is always 0 while for trustful strategies the payoffs are

$$\begin{aligned} f_1 &= x_1 R + y_1 S \\ g_1 &= x_1 T + y_1 P, \end{aligned} \tag{2}$$

where f_1, g_1 are the expected payoffs and x_1, y_1 are the frequencies of trustful cooperators and trustful defectors, respectfully. We will use subscript 1 to denote trustful players, and we reserve subscript 0 to denote suspicious players. The subscripts can be thought of representing the probability of accepting unknown opponents.

The evolutionary dynamics of this model can be solved fully analytically (SI) and the results are depicted in Fig. 2. In Fig. 2(a) where $0 < P$, all trajectories approach trustful defection. In Fig. 2 (b) where $P = 0$, all trajectories approach the line of equilibria spanned by suspicious strategies and trustful defection, and in Fig. 2(c) where $P < 0$, all trajectories approach suspicious strategies. Note that in the last case the boundary is a heteroclinic cycle. This model was analysed in the context of public goods game by Hauert et al. (2002a; 2002b).

If players have perfect information about the type of the opponent $q = 1$, it is nonsensical to distinguish between suspicious and trustful strategies as all opponents are known. In both non-trivial cases $S < 0 \leq P < R < T$ and $S < P < 0 < R < T$ cooperators will only accept interactions with other cooperators, while defectors will accept defectors only if $0 < P$. For all payoffs no games between defectors and cooperators are played. The analysis of the evolutionary dynamics is straightforward. If $0 < P$ cooperation and defection are both locally attracting states separated by an unstable equilibrium (Fig. 2(d)), and if $P \leq 0$ cooperation is globally attracting (Fig. 2(e)).

3.2. Games with partial information

In this section we consider models with partial information $0 < q < 1$. The first model we analyze is where opting out of interactions yields no benefits nor costs to the player and so $P = 0$. We analyze this case first because of its simple evolutionary dynamics and because it contains the donation game, a version of prisoners dilemma that has a central role in the literature of the evolution of cooperation (Sigmund, 2010).

3.2.1. Opting out yields no benefits nor costs

In this section we assume that opting out yields players the same payoff as mutual defection, i.e. $P = 0$. In such a case, defectors should always accept unknown players since accepting a game guarantees them a payoff that is at least 0 ($\leq P, T$). Suspicious defection is therefore not a rational strategy and will not be considered. Cooperators, however, may want to accept or opt out of an interaction with an unknown player: if the opponent is likely to be a cooperator, accepting is more beneficial than opting out $0 < R$, but if the opponent is likely to be a defector it is better to opt out $S < 0$. We thus consider three strategies, trustful cooperators who accept a known cooperator and an unknown opponent but reject a known defector, suspicious cooperators who accept a known cooperator but reject everyone else, and trustful defectors who accept all opponents.

To investigate the evolutionary dynamics (1) we calculate the expected payoffs for each strategy

$$\begin{aligned} f_0 &= \left(x_0 q^2 + x_1 q\right) R \\ f_1 &= (x_0 q + x_1) R + y_1 (1 - q) S \\ g_1 &= x_1 (1 - q) T, \end{aligned} \tag{3}$$

where similarly to previous section f_0, f_1, g_1 are the expected payoffs and x_0, x_1, y_1 are the frequencies of suspicious cooperators, trustful cooperators and trustful defectors, respectfully.

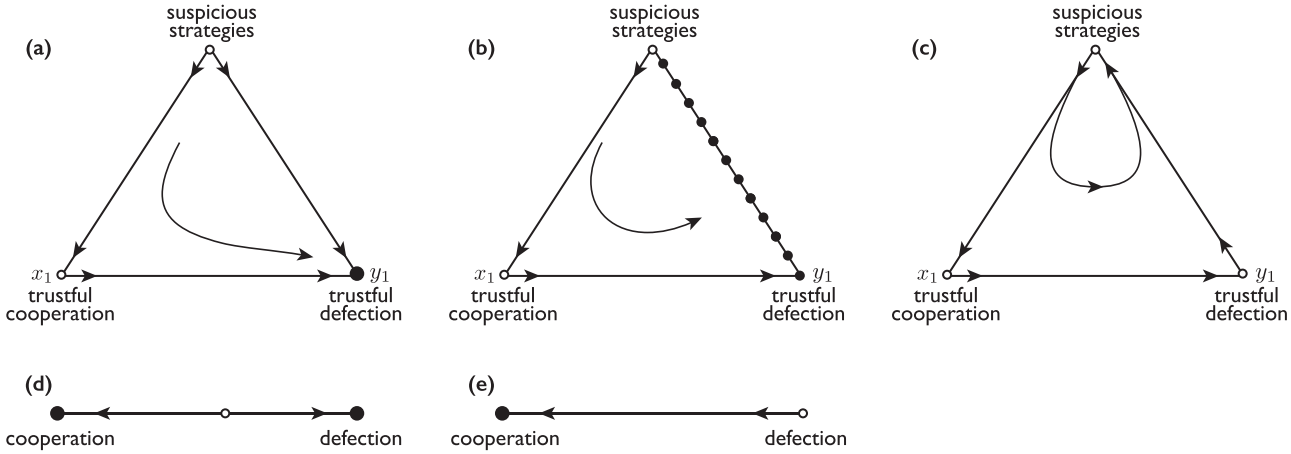

Fig. 2. Top row: evolutionary dynamics (1) for a model (2) with zero information $q = 0$. In (a) $0 < P$ all trajectories approach trustful defection (b) $P = 0$ all trajectories approach the line of equilibria spanned by suspicious strategies and trustful defection (c) $P < 0$ all trajectories approach suspicious strategies. Note that the boundary is a heteroclinic cycle. Bottom row: evolutionary dynamics (1) for a model with perfect information $q = 1$. In (d) $0 < P$ cooperation and defection are locally attracting, separated by an unstable equilibrium. In (e) $P \leq 0$ all trajectories approach cooperation.

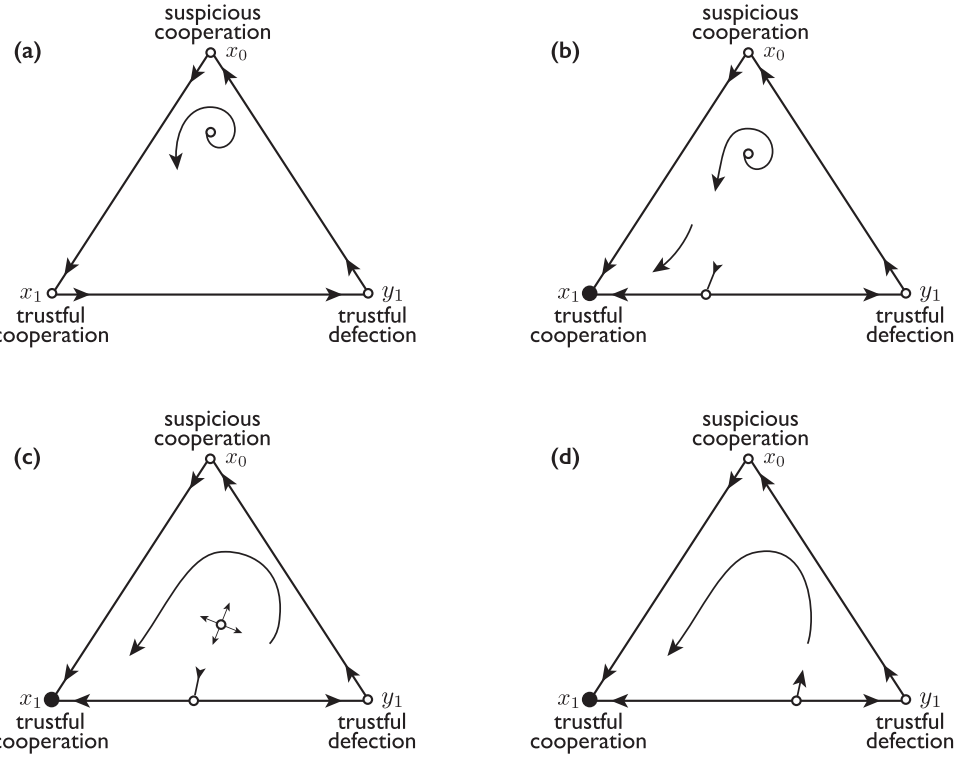

Fig. 3. Evolutionary dynamics (1) for a model (3) where opting out is not costly nor beneficial $P = 0$. The parameter values are **(a)** $0 < q < \frac{T-R}{T}$ **(b)** $\frac{T-R}{T} < q < \frac{T}{R(1+\frac{R}{4T})+T}$ **(c)** $\frac{T}{R(1+\frac{R}{4T})+T} < q < \frac{T}{R+T}$ **(d)** $\frac{T}{R+T} < q < 1$. In each panel in the top node all players are suspicious cooperators ($x_0 = 1$), in the bottom left node all players are trustful cooperators ($x_1 = 1$) and in the bottom right node all players are trustful defectors ($y_1 = 1$). The analytical expressions for the dimorphic and trimorphic equilibria, and their stability conditions, are given in the SI. There are two qualitatively different evolutionary trajectories: In panel (a) $0 < q < \frac{T-R}{T}$ every trajectory approaches the rock-paper-scissors cycle of trustful cooperation, trustful defection and suspicious cooperation, and in panels (b)–(d) $\frac{T-R}{T} < q < 1$ all trajectories converge to a fully trustful cooperation.

The evolutionary dynamics (1) with the expected payoffs given in (3) can be analysed fully analytically (see SI) and the results are depicted in Fig. 3. In Fig. 3(a), where $0 < q < \frac{T-R}{T}$, all trivial equilibria are saddles and because the (strictly) interior trimorphic equilibrium (x_0, x_1, y_1) is an unstable spiral all trajectories approach the heteroclinic cycle of trustful cooperation, trustful defection and suspicious cooperation (see SI for the exact expression of the interior trimorphic equilibrium and the stability analysis). In Fig. 3(b) where $\frac{T-R}{T} < q < \frac{T}{R(1+\frac{R}{4T})+T}$, trustful cooperation turns into a stable equilibrium, and so all trajectories approach the equilibrium of trustful cooperation. In Fig. 3(c) where $\frac{T}{R(1+\frac{R}{4T})+T} < q < \frac{T}{R+T}$, the interior trimorphic equilibrium (x_0, x_1, y_1) changes from an unstable spiral to an unstable node, and in Fig. 3 (d) where $\frac{T}{R+T} < q < 1$, the trimorphic equilibrium (x_0, x_1, y_1) exits the interior. In both cases all trajectories approach the equilibrium of trustful cooperation. We remark that in the limiting cases where q approaches 0 or 1 we recover the model with zero $q = 0$ and perfect information $q = 1$, respectfully: as q approaches 0 the trimorphic equilibrium (x_0, x_1, y_1) approaches the equilibrium of suspicious cooperation x_0 and the line spanned by suspi-

cious cooperators x_0 and trustful defectors y_1 turns into a line of equilibria (Fig. 2(b)), and as q approaches 1 the unstable dimorphic equilibrium (x_1, y_1) approaches the equilibrium of trustful defection y_1 and so all trajectories approach the equilibrium of trustful cooperation.

We have obtained two qualitatively different evolutionary outcomes. First, when $0 < q < \frac{T-R}{T}$, the evolutionary dynamics approaches a heteroclinic rock-paper-scissors cycle of trustful cooperation, trustful defection and suspicious cooperation (Fig. 3(a)). This is because for lower values of q most encounters are between unknown players. Therefore (i) almost all games between trustful defectors and trustful cooperators are accepted, and the situation is (almost) identical to the donation game with obligatory interactions where trustful defection beats trustful cooperation (ii) when trustful cooperators are absent both suspicious cooperators and trustful defectors play only amongst themselves, and because cooperative interaction yields higher payoff than defective interactions suspicious cooperators beat trustful defectors (iii) if most players are cooperators, trustful cooperators beat suspicious cooperators because trustful cooperators play more cooperative games by accepting unknown, and therefore cooperative, opponents.

Second, when $\frac{T-R}{T} < q < 1$, the evolutionary outcome is a population of trustful cooperation, independently of the initial (strictly positive) frequency distribution of strategies (Figs. 3(b)–(d)). Trustful cooperation is an ESS because for higher values of q a population of trustful cooperators efficiently refuse defective opponents. This implies that trajectories nearby converge to a fully trustful cooperation. The global convergence is due to the existence of suspicious cooperators as they can invade a population of defectors, and then be eventually replaced by trustful cooperators.

We remark that a similar ESS condition was derived in Nowak and Sigmund (1998a), Nowak and Sigmund (1998b), Suzuki and Toquenaga (2005) and Ghang and Nowak (2015). There are however two notable differences. Firstly, the condition given in the previous work was derived for a donation game stating that cooperation is an ESS if the probability of knowing the type of the opponent q is greater than the cost to benefit ratio of cooperation. However, our model is derived for the general prisoners dilemma allowing us to make a distinction between the cost of cooperation $P - S$ and the benefit of defection $T - R$ (in the donation game they are equal). The interpretation of the ESS condition then becomes a ratio between the benefit of defection $T - R$, rather than cost of cooperation, and a payoff value which is the difference between unknown and known defectors encountering a trustful cooperator, i.e. T (recall the reinterpretation of the payoff values). Secondly, but more importantly, our condition implies global convergence to trustful cooperation. This is a consequence of allowing decision rules that are optimal when trustful behaviour is not, and therefore, when population consist mainly of defectors, suspicious behaviour becomes the outcompeting social norm which eventually enables the dominance of trustful cooperation.

3.2.2. Opting out is costly

Lets now suppose that players who opt out are strictly worse off than players who mutually defect $S < 0 < P$. Because defectors should accept every interaction whenever $0 \leq P$, the strategies under consideration are identical to the previous model ($P = 0$). The expected payoffs are

$$\begin{aligned} f_0 &= \left(x_0 q^2 + x_1 q\right) R \\ f_1 &= (x_0 q + x_1) R + y_1 (1 - q) S \\ g_1 &= x_1 (1 - q) T + y_1 P. \end{aligned} \tag{4}$$

The evolutionary dynamics (1) with the expected payoffs given in (4) can be analysed fully analytically (see SI for detailed analysis) and we summarise the results in Fig. 4.

In contrast with the previous model with $P = 0$, a trimorphic equilibrium (x_0, x_1, y_1) enters the interior of the state space only for intermediate values of q and only if $\frac{P}{-S} < 1$ holds. Consequently, there are three cases to consider that depend on whether the trimorphic equilibrium enters the interior of the state space, and if it does, whether at the time of entry the equilibrium of trustful cooperation is stable or not. In the first case the trimorphic equilibrium (x_0, x_1, y_1) enters the interior while trustful cooperation is an unstable equilibrium $0 < \frac{P}{-S} < \frac{T-R}{T}$ (Fig. 4, top row (a1)–(d1)). In (a1) $0 < q < \frac{P}{-S}$ the trimorphic equilibrium (x_0, x_1, y_1) is in the exterior of the state space and the only stable equilibrium is the equilibrium of trustful defection y_1. In (b1) $\frac{P}{-S} < q < \frac{T-R}{T}$ the unstable trimorphic equilibrium (x_0, x_1, y_1) enters the interior, and in (c1) $\frac{T-R}{T} < q < \frac{PR-ST}{-S(R+T)}$ the equilibrium of trustful cooperation x_1 becomes stable. In (d1) $\frac{PR-ST}{-S(R+T)} < q < 1$ the trimorphic equilibrium (x_0, x_1, y_1) leaves the interior. We have that in (a1)–(b1) all evolutionary trajectories approach the equilibrium of trustful defection y_1 (globally convergent ESS), and in (c1)–(d1) it depends on the initial frequency distribution of strategies whether the evolutionary trajectories approach the equilibrium of trustful cooperation x_1 or trustful defection y_1 (both locally convergent ESS).

In the second case trustful cooperation is stable as the trimorphic equilibrium (x_0, x_1, y_1) enters the interior $0 < \frac{T-R}{T} < \frac{P}{-S} < 1$ (Fig. 4, middle row (a2)–(d2)). In (a2) $0 < q < \frac{T-R}{T}$ the trimorphic equilibrium (x_0, x_1, y_1) is in the exterior of the state space and the only stable equilibrium is the equilibrium of trustful defection y_1. In (b2) $\frac{T-R}{T} < q < \frac{P}{-S}$ the equilibrium of trustful cooperation becomes stable, in (c2) $\frac{P}{-S} < q < \frac{PR-ST}{-S(R+T)}$ the unstable trimorphic equilibrium (x_0, x_1, y_1) enters the interior and in (d2) $\frac{PR-ST}{-S(R+T)} < q < 1$ the unstable trimorphic equilibrium (x_0, x_1, y_1) leaves the interior. We have that in (a2) all trajectories approach the equilibrium of trustful defection y_1 and in (b2)–(d2) it depends on the initial frequency distribution of strategies whether the trajectories approach the equilibrium of trustful cooperation x_1 or trustful defection y_1. In the third case the trimorphic equilibrium never enters the interior $1 < \frac{P}{-S}$ (Fig. 4, bottom row (a3)–(b3)). In (a3) $0 < q < \frac{T-R}{T}$ the only stable equilibrium is the equilibrium of trustful defection y_1 and so all trajectories approach trustful defection and in (b3) $\frac{T-R}{T} < q < 1$ the equilibrium of trustful cooperation becomes stable and so depending on the initial frequency distribution of strategies all the trajectories approach the equilibrium of trustful cooperation x_1 or trustful defection y_1. We remark that as q approaches 0 or 1 this model simplifies to the model with zero $q = 0$ (Fig. 2a) and perfect information $q = 1$ (Fig. 2d), respectfully.

We observe that in this model trustful defection is an ESS for all values of q. This is because opting out is costly $0 < P$ and so both trustful and suspicious cooperators are at a disadvantage for sufficiently high frequency of defectors. This means that all trajectories converge to a fully defective population whenever trustful defection is the only stable equilibrium $0 < q < \frac{T-R}{T}$. When $\frac{T-R}{T} < q < 1$ trustful cooperation is also an ESS, but contrary to the previous model ($P = 0$) it is not a globally convergent ESS. However, the basin of attraction increases with q and for large q only trajectories close to full defection are unable to reach the ESS of trustful cooperation.

3.2.3. Opting out is beneficial

In this section we suppose that opting out yields a strictly greater payoff than mutual defection $P < 0 < R$. In contrast to the previous two cases, defectors ought to avoid each other and so in addition to trustful cooperators, trustful defectors and suspicious cooperators we must also consider suspicious defectors, having in total four strategies. Note that since in this model mutual defection is worse than opting out, defective strategies will reject known de-

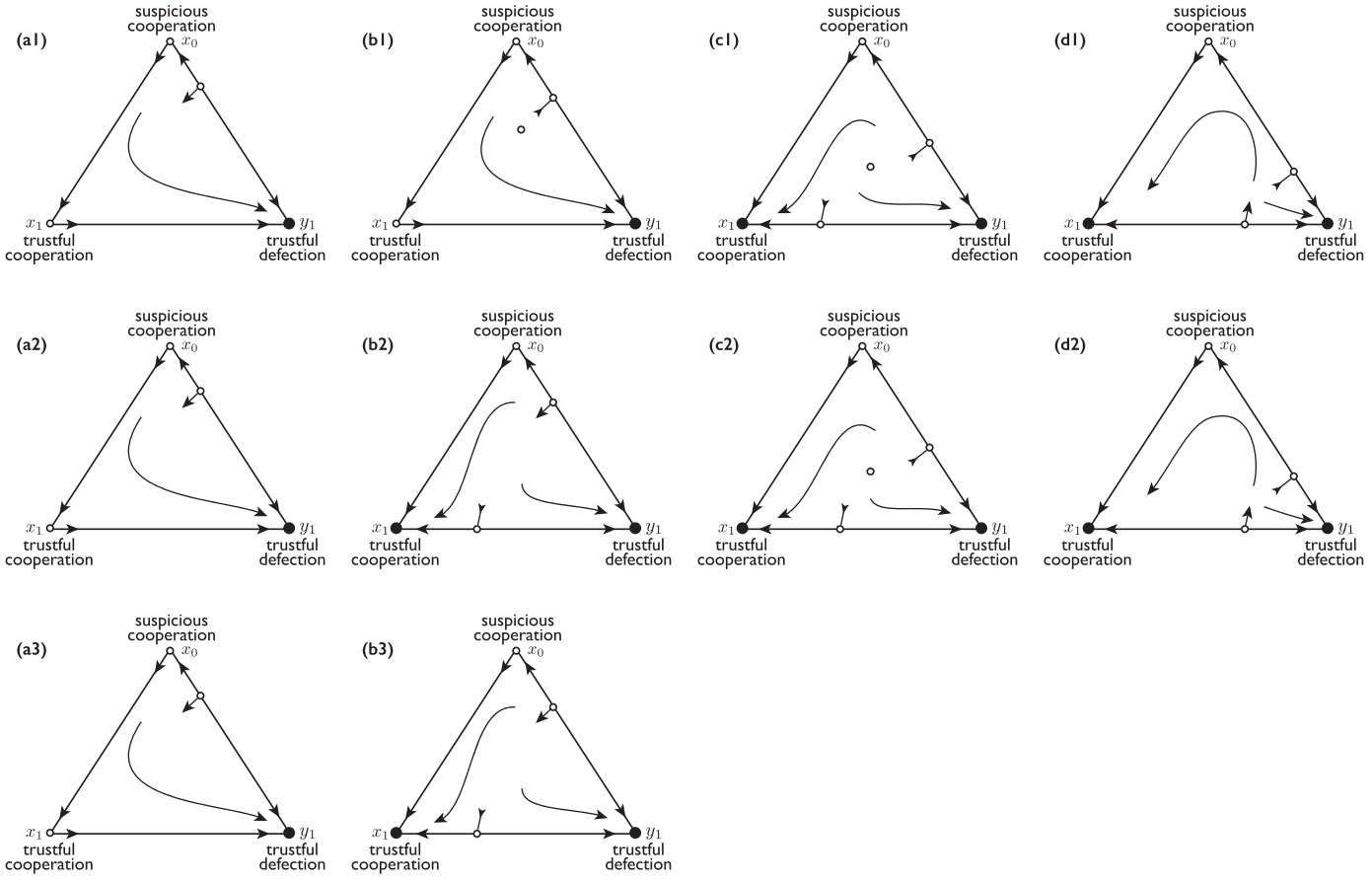

Fig. 4. Evolutionary dynamics (1) for a model (4) where opting out is costly $S<0<P$. We distinguish three cases (a1)–(d1), (a2)–(d2) and (a3)–(b3), depending on the relationship between the trimorphic equilibrium (x_0, x_1, y_1) and the equilibrium of trustful cooperation (see the main text). The parameter values are **(a1)** $0<q<\frac{P}{-S}$ **(b1)** $\frac{P}{-S}<q<\frac{T-R}{T}$ **(c1)** $\frac{T-R}{T}<q<\frac{PR-ST}{-S(R+T)}$ **(d1)** $\frac{PR-ST}{-S(R+T)}<q<1$, **(a2)** $0<q<\frac{T-R}{T}$ **(b2)** $\frac{T-R}{T}<q<\frac{P}{-S}$ **(c2)** $\frac{P}{-S}<q<\frac{PR-ST}{-S(R+T)}$ **(d2)** $\frac{PR-ST}{-S(R+T)}<q<1$, **(a3)** $0<q<\frac{T-R}{T}$ **(b3)** $\frac{T-R}{T}<q<1$. Notation is identical to Fig. 3. There are two qualitatively different evolutionary outcomes: in panels where $0<q<\frac{T-R}{T}$ all trajectories approach trustful defection, and in panels where $\frac{T-R}{T}<q<1$ all trajectories approach either trustful defection or trustful cooperation depending on the initial frequency distribution. See SI for a detailed analysis.

fectors. The expected payoffs are

$$\begin{aligned} f_0 &= \left(x_0 q^2 + x_1 q\right)R \\ f_1 &= (x_0 q + x_1)R + (y_0 q + y_1)(1-q)S \\ g_0 &= x_1 q(1-q)T \\ g_1 &= x_1(1-q)T + y_1(1-q)^2 P, \end{aligned} \tag{5}$$

where y_0 is the frequency and g_0 the expected payoff of suspicious defectors. The evolutionary dynamics (1) with the expected payoffs given in (5) can be analysed analytically, except for intermediate values of q where we couldn't determine which of the two, when $T<4R$, or three, when $4R\leq T$, possible heteroclinic cycles evolutionary trajectories approach to (see below for the precise condition; a more detailed analysis is in SI). Fig. 5 summarizes the results for the case $T<4R$ and Fig. 6 summarizes the case $4R\leq T$.

The threshold values at which we transition between panels in Figs. 5 and 6 are

$$q^{\text{stab.}}_{x_0x_1y_1} = \frac{1}{-2P(T-R)}\left[-2P(T-R)-SR-\sqrt{R^2S^2+4SPRT-4SPR^2}\right] \tag{6}$$

$$q_0 = \frac{T-R}{T} = q^{\text{enter}}_{x_1y_1} = q^{\text{exit}}_{x_0x_1y_0} \tag{7}$$

$$q^{\text{exit}}_{x_0x_1y_1} = \frac{-1}{2PR}[S(R+T)-2PR+\sqrt{S^2(R+T)^2-4PSR^2}] \tag{8}$$

$$q^{\text{entry}}_{x_1y_0y_1} = \frac{1}{-2PT}[T(S-P)+\sqrt{-4P^2RT+T^2(P+S)^2}] \tag{9}$$

where $q^{\text{stab.}}_{x_0x_1y_1}<q_0<q^{\text{exit}}_{x_0x_1y_1}<q^{\text{entry}}_{x_1y_0y_1}$ for all payoff values S, P, R, T. In Fig. 6 we need additional thresholds

$$q^{\text{exit}}_{x_1y_0} = \frac{1}{2}-\frac{1}{2}\sqrt{1-4\frac{R}{T}} \tag{10}$$

$$q^{\text{entry}}_{x_1y_0} = \frac{1}{2}+\frac{1}{2}\sqrt{1-4\frac{R}{T}}, \tag{11}$$

where $q^{\text{exit}}_{x_1y_0}<q^{\text{entry}}_{x_1y_0}<q_0$ for all R, T. However, the relative order between the thresholds $q^{\text{exit}}_{x_1y_0}$, $q^{\text{entry}}_{x_1y_0}$ and $q^{\text{stab.}}_{x_0x_1y_1}$ depends on S, P, R, T.

Let us first consider the case $T<4R$ (Fig. 5). In panel (a) $0<q<q^{\text{stab.}}_{x_0x_1y_1}$ there exists a stable trimorphic equilibrium (x_0, x_1, y_1). In panel (b) $q^{\text{stab.}}_{x_0x_1y_1}<q<q_0$ the trimorphic equilibrium (x_0, x_1, y_1) becomes unstable and there are no stable equilibria in the system. In panel (c) $q_0<q<q^{\text{exit}}_{x_0x_1y_1}$ the dimorphic equilibrium (x_1, y_1) enters the interior and trustful cooperation x_1 becomes stable. In panel (d) $q^{\text{exit}}_{x_0x_1y_1}<q<q^{\text{entry}}_{x_1y_0y_1}$ the trimorphic equilibrium (x_0, x_1, y_1) exits the interior by passing through the dimorphic equilibrium (x_1, y_1), and in panel (e) $q^{\text{entry}}_{x_1y_0y_1}<q<1$ an unstable trimorphic equilibrium (x_1, y_0, y_1) enters the interior by passing through

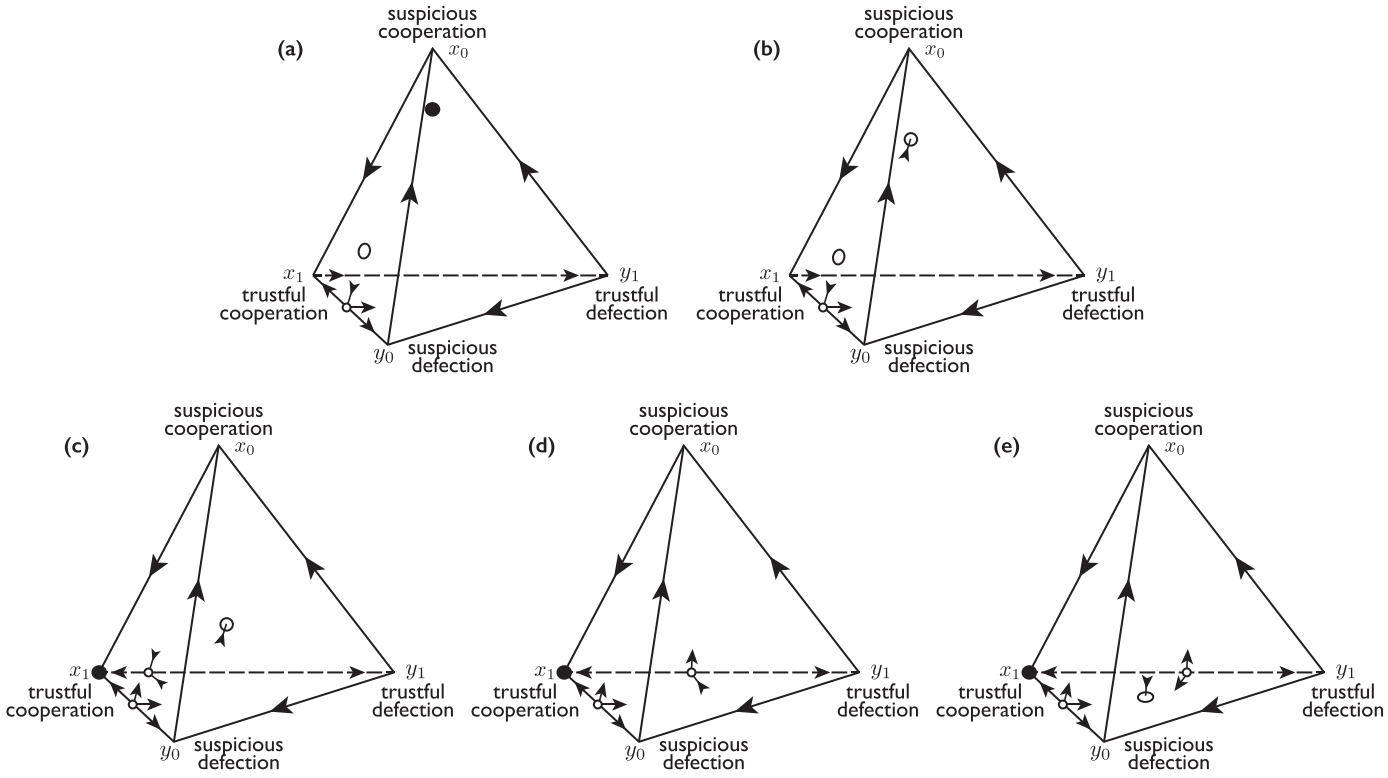

Fig. 5. Evolutionary dynamics (1) for a model (5) where opting out is beneficial $P<0<R$, and where $T<4R$: **(a)** $0<q<q^{\text{stab.}}_{x_0x_1y_1}$ **(b)** $q^{\text{stab.}}_{x_0x_1y_1}<q<q_0$ **(c)** $q_0<q<q^{\text{exit}}_{x_0x_1y_1}$ **(d)** $q^{\text{exit}}_{x_0x_1y_1}<q<q^{\text{entry}}_{x_1y_0y_1}$ **(e)** $q^{\text{entry}}_{x_1y_0y_1}<q<1$. The filled circles are stable equilibria, i.e. all the eigenvalues are negative (see SI for details). For simplicity no arrows are drawn for the trimorphic equilibria, unless the equilibrium is an unstable equilibrium but also has negative eigenvalues in which case the stable direction(s) is drawn. There are three different evolutionary outcomes. **1**. All trajectories approach the equilibrium of suspicious cooperation, trustful cooperation and trustful defection (x_0, x_1, y_0) (panel (a)). **2**. All trajectories approach one of the two heteroclinic cycles, either $x_0\to x_1\to y_1$ or $x_0\to x_1\to y_1\to y_0$. Numerical investigation shows it is the first one (panel (b)). **3**. All trajectories approach the equilibrium of trustful cooperation x_1 (panels (c)-(d)).

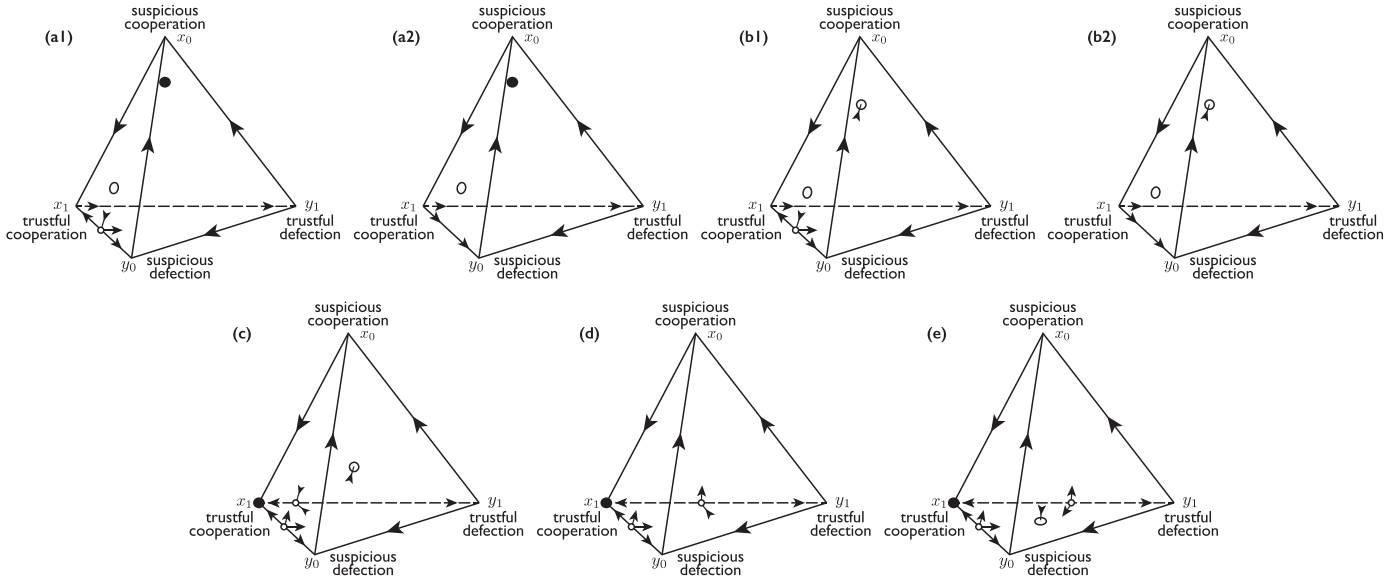

Fig. 6. Evolutionary dynamics (1) for a model (5) where opting out is beneficial $P<0<R$, and where $4R\le T$: In contrast to the case in Fig. 5, the unstable equilibrium (x_1, y_0) exits the interior for $q^{\text{exit}}_{x_1y_0}<q<q^{\text{entry}}_{x_1y_0}$. We distinguish three cases based on the order in which we transition between the panels when q increases. For the case **(i)** $q^{\text{stab.}}_{x_0x_1y_1}<q^{\text{exit}}_{x_1y_0}$, we transition between (a1), (b1), (b2), (b1), after which continue to (c), (d) and (e) **(ii)** $q^{\text{exit}}_{x_1y_0}<q^{\text{stab.}}_{x_0x_1y_1}<q^{\text{entry}}_{x_1y_0}$ we transition between (a1), (a2), (b2), (b1), after which continue to (c), (d) and (e), and **(iii)** $q^{\text{entry}}_{x_1y_0}<q^{\text{stab.}}_{x_0x_1y_1}$ we transition between (a1), (a2), (a1), (b1), after which continue to (c), (d) and (e). Similarly to Fig. 5 we have **(a1a2)** $0<q<q^{\text{stab.}}_{x_0x_1y_1}$ **(b1b2)** $q^{\text{stab.}}_{x_0x_1y_1}<q<q_0$ **(c)** $q_0<q<q^{\text{exit}}_{x_0x_1y_1}$ **(d)** $q^{\text{exit}}_{x_0x_1y_1}<q<q^{\text{entry}}_{x_1y_0y_1}$ **(e)** $q^{\text{entry}}_{x_1y_0y_1}<q<1$. Notation is identical to Fig. 5. There are three different evolutionary outcomes: **1**. All trajectories approach the equilibrium of suspicious cooperation, trustful cooperation and trustful defection (x_0, x_1, y_0) (panels (a1, a2)). **2**. All trajectories approach one of the three heteroclinic cycles, either $x_0\to x_1\to y_1$ or $x_0\to x_1\to y_1\to y_0$ (panels b1,b2), or an additional cycle $x_0\to x_1\to y_0$ which is possible only in panel (b2). Numerical investigation shows it is the first one. **3**. All trajectories approach the equilibrium of trustful cooperation x_1 (panels (c)-(d)).

the dimorphic equilibrium (x_1, y_0). Because there are no interior 4−morphic equilibria (see SI) all evolutionary trajectories approach the boundary of the state space. As a consequence we get that in panel (a) all evolutionary trajectories approach the stable coexistence of suspicious cooperation, trustful cooperation and trustful defection at the equilibrium (x_0, x_1, y_1). In panel (b) all evolutionary trajectories approach one of the two heteroclinic cycles, either the cycle between suspicious cooperation, trustful cooperation and trustful defection ($x_0 \to x_1 \to y_1$) or the cycle between suspicious cooperation, trustful cooperation, trustful defection and suspicious defection ($x_0 \to x_1 \to y_1 \to y_0$). Our numerical investigation indicates it is the cycle $x_0 \to x_1 \to y_1$. Finally, in panels (c)–(e) all evolutionary trajectories approach trustful cooperation x_1.

In Fig. 6, where $4R \le T$, the phase planes are similar to the previous case except that the dimorphic unstable equilibrium (x_1, y_0) exits the interior for $q^{\text{exit}}_{x_1y_0} < q < q^{\text{entry}}_{x_1y_0}$. We need to distinguish three cases based on the order in which we transition between the panels when q increases. In the first case (i) $q^{\text{stab.}}_{x_0x_1y_1} < q^{\text{exit}}_{x_1y_0}$, we transition between (a1), (b1), (b2), (b1), after which we continue to (c), (d) and (e). In the second case (ii) $q^{\text{exit}}_{x_1y_0} < q^{\text{stab.}}_{x_0x_1y_1} < q^{\text{entry}}_{x_1y_0}$ we transition between (a1), (a2), (b2), (b1), after which we continue to (c), (d) and (e), and (iii) $q^{\text{entry}}_{x_1y_0} < q^{\text{stab.}}_{x_0x_1y_1}$ we transition between (a1), (a2), (a1), (b1), after which we continue to (c), (d) and (e). Otherwise the threshold values for which we transition between panels are similar to Fig. 5. An important consequence of the dimorphic equilibrium exiting the interior is that in panel (b2) evolutionary trajectories may approach an additional heteroclinic cycle of suspicious cooperation, trustful cooperation and suspicious defection ($x_0 \to x_1 \to y_0$). However, our numerical investigation indicates all trajectories approach the cycle $x_0 \to x_1 \to y_1$. We remark that as q approaches 0 or 1 this model simplifies to the model with no $q = 0$ (Fig. 2c) and perfect information $q = 1$ (Fig. 2e), respectfully. As q approaches 0 then the globally stable trimorphic equilibrium (x_0, x_1, y_1) approaches the equilibrium of suspicious cooperation x_0 and when q approaches 1 then the unstable dimorphic equilibrium (x_1, y_1) approaches y_1 and so all trajectories approach trustful cooperation.

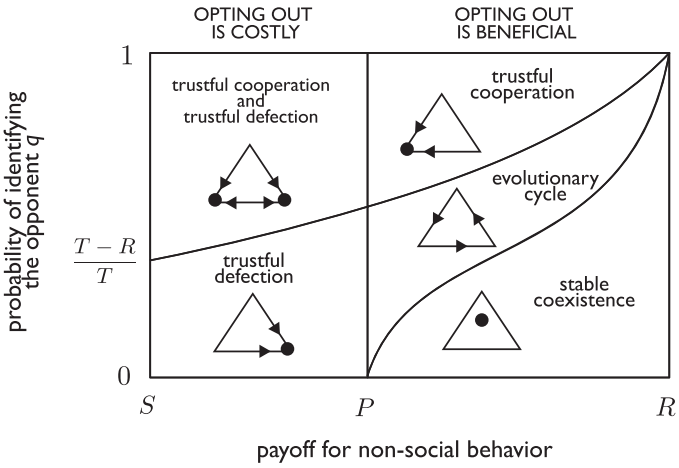

Fig. 7. Summary of the results. On the vertical axis is the probability of knowing the type of the opponent q, and on the horizontal axis is the payoff for non-social behaviour 0 (opting out). The vertical line in the middle represents the non-generic case $P = 0$, while on the left of the vertical line opting out is costly $S < 0 < P$ and on the right opting out is beneficial $P < 0 < R$. In each area we draw a triangle that represents the phase plane for the parameter values in the area, such that in each triangle in the bottom left corner all players are trustful cooperators x_1, in the bottom right corner all players are trustful defectors y_1 and the upper corner all players are suspicious cooperators x_1. Trustful cooperation is an ESS above the curve $q = \frac{T-R}{T}$ (the upper curve) and trustful defection is an ESS whenever $S < 0 < P$. Thus for $S < 0 < P$ and $0 \le q \le \frac{T-R}{T}$ all trajectories approach trustful defection, for $P \le 0 < R$ and $\frac{T-R}{T} < q \le 1$ all trajectories approach trustful cooperation and for $S < 0 < P$ and $\frac{T-R}{T} < q \le 1$ all trajectories approach either trustful defection or trustful cooperation depending on the initial frequency distribution. For $P \le 0 < R$ and $q^{\text{stab.}}_{x_0x_1y_1} < q \le \frac{T-R}{T}$, where $q = q^{\text{stab.}}_{x_0x_1y_1}$ is the bottom curve (see the exact expression in (6)), all trajectories approach the rock-paper-scissors cycle of suspicious cooperation, trustful cooperation and trustful defection (numerical result). For $P < 0 < R$ below the curve $q = q^{\text{stab.}}_{x_0x_1y_1}$ all trajectories approach the stable coexistence of suspicious cooperation, trustful cooperation and trustful defection.

4. Discussion

In this paper we introduced an evolutionary game theoretic model where individuals encounter each other at random, but have the option to opt out of interactions based on partial information about their encountered opponents. With a fixed probability, individuals are assumed to know whether the opponent is a cooperator or defector. This simple formulation allowed us to solve the model of prisoners dilemma with optional interactions fully analytically, with the exception of a specific parameter region where we were not able to determine which of the three or four heteroclinic cycles evolutionary trajectories approach to (see below).

The results of our paper are summarised in Fig. 7. First, we find that if the probability of identifying the type of the opponent is sufficiently high, $\frac{T-R}{T} < q \le 1$, then trustful cooperation is an ESS (similar condition was derived in Nowak and Sigmund, 1998a; 1998b; Suzuki and Toquenaga, 2005; Ghang and Nowak, 2015). Interestingly, and in contrast with previous findings, if opting out is at least as beneficial as mutual defection ($P \le 0$), then trustful cooperation is a globally convergent ESS, i.e. trustful cooperation is reached from any initial frequency distribution of strategies (Fig. 7). In particular, even an (almost) entirely defective population will be replaced by trustful cooperators.

Secondly, we find that if the probability of knowing the type of the opponent is $0 < q \le \frac{T-R}{T}$, and opting out is at least as beneficial as mutual defection ($P \le 0$), then all evolutionary trajectories approach one of the three heteroclinic cycles given in model (5) (area denoted "evolutionary cycle" in Fig. 7). Numerical investigation (in the case $P = 0$ analytical analysis) indicates that all trajectories approach the cycle of suspicious cooperation, trustful cooperation and trustful defection. Thirdly, if opting out is strictly worse than mutual defection ($S < 0 < P$) then trustful defection is always an ESS, either a locally convergent $\frac{T-R}{T} < q \le 1$ (area denoted "trustful defection and trustful cooperation" in Fig. 7) or globally convergent ESS $0 \le q \le \frac{T-R}{T}$ (red area denoted "trustful defection" in Fig. 7). Lastly, if opting out is strictly beneficial ($P < 0 < R$), then for $0 \le q < q^{\text{stab.}}_{x_0x_1y_1}$, trustful cooperators, trustful defectors and suspicious cooperators coexist at a globally stable equilibrium (see model (5) for the exact condition; area denoted "stable coexistence" in Fig. 7). Note that suspicious defectors are always (eventually) selected against and thus eradicated from the population. We remark that the models with zero $q = 0$ and perfect information $q = 1$ are aligned with the Fig. 7.

Our model can be extended in a straightforward manner to several interesting directions. One possibility is to consider a multiplayer version of our model where each player has partial information about other players in the group. Here, a group of players may find themselves in a situation where only a fraction of players want to opt out while others would wish to continue the game, ultimately requiring to include more complex decision-rules as the group size increases. Another possibility is to allow errors in perception or execution of strategies (Molander, 1985; Sigmund, 2010). This scenario would also require updating our current strategies as even trustful individuals should doubt the truthfulness of the observed type (errors in perception) or should be suspicious of the future action of the opponent (errors in execution). Yet another possibility is to consider a game where players don't have the option of opting out if the opponent wants to interact. This case may apply for example in mating systems

with forced copulations (Verrell, 1998). However, the assumption of forced interactions may be suited better for games other than prisoners dilemma where we suspect its effect on the dynamics becomes trivial. This is because in prisoners dilemma the preference for opponents is unidirectional, and so the preferred cooperative players would be forced into harmful partnerships, consequently lowering the level of cooperation. Finally, instead of pure-decision rules assumed in this paper a mixed decision could be used where accepting an unknown opponent happens with some probability. This set-up could be used, for example, to investigate the gradual evolution of trust in fully suspicious populations.

To conclude, our simple mathematically tractable evolutionary model with optional interactions, a model that can be readily extended to games other than prisoners dilemma, shows that the option of non-social behaviour facilitates the emergence of cooperative behaviour. Interestingly, the option of non-sociality facilitates not only stable cooperative populations but also trustful behaviour that accepts interactions with potentially harmful players.

Acknowledgments

The research leading to these results has received funding from the People Programme (Marie Curie Actions) of the European Union's Seventh Framework Programme (FP7/2007–2013) under REAGrant Agreement No. 291734, the Austrian Science Fund (FWF) S11407-N23 (RiSE/SHiNE), the ERCStart Grant (279307: Graph Games) and the John Templeton Foundation.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at 10.1016/j.jtbi.2017.08.025.

References

Andersson, M., Simmons, L.W., 2006. Sexual selection and mate choice. Trends Ecol. Evol. 21 (6), 296–302.
Batali, J., Kitcher, P., 1995. Evolution of altriusm in optional and compulsory games. J. Theor. Biol. 175 (2), 161–171.
Brandt, H., Hauert, C., Sigmund, K., 2006. Punishing and abstaining for public goods. Proc. Natl. Acad.Sci. 103 (2), 495–497.
Cardinot, M., Gibbons, M., O'Riordan, C., Griffith, J., 2016. Simulation of an optional strategy in the prisoner's dilemma in spatial and non-spatial environments. In: International Conference on Simulation of Adaptive Behavior. Springer, pp. 145–156.
Castro, L., Toro, M.A., 2008. Iterated prisoners dilemma in an asocial world dominated by loners, not by defectors. Theor. Popul. Biol. 74 (1), 1–5.
DeSteno, D., Breazeal, C., Frank, R.H., Pizarro, D., Baumann, J., Dickens, L., Lee, J.J., 2012. Detecting the trustworthiness of novel partners in economic exchange. Psychol. Sci. 0956797612448793.
Fowler, J.H., 2005. Altruistic punishment and the origin of cooperation. Proc. Natl. Acad. Sci. U.S.A. 102 (19), 7047–7049.
Frank, R.H., Gilovich, T., Regan, D.T., 1993. The evolution of one-shot cooperation: an experiment. Ethology Sociobiol. 14 (4), 247–256.
Fu, F., Hauert, C., Nowak, M.A., Wang, L., 2008. Reputation-based partner choice promotes cooperation in social networks. Phys. Rev. E 78 (2), 026117.
Fujiwara-Greve, T., Okuno-Fujiwara, M., 2009. Voluntarily separable repeated prisoner's dilemma. Rev. Econ. Stud. 76 (3), 993–1021. doi:10.1111/j.1467-937X.2009.00539.x.
Ghang, W., Nowak, M.A., 2015. Indirect reciprocity with optional interactions. J. Theor. Biol. 365, 1–11.
Hauert, C., De Monte, S., Hofbauer, J., Sigmund, K., 2002a. Replicator dynamics for optional public good games. J. Theor. Biol. 218 (2), 187–194.
Hauert, C., De Monte, S., Hofbauer, J., Sigmund, K., 2002b. Volunteering as red queen mechanism for cooperation in public goods games. Science 296 (5570), 1129–1132.
Hauert, C., Traulsen, A., Brandt, H., Nowak, M.A., Sigmund, K., 2007. Via freedom to coercion: the emergence of costly punishment. Science 316 (5833), 1905–1907.
Hruschka, D.J., Henrich, J., 2006. Friendship, cliquishness, and the emergence of cooperation. J. Theor. Biol. 239 (1), 1–15.
Iwasa, Y., Pomiankowski, A., Nee, S., 1991. The evolution of costly mate preferences II. the'handicap'principle. Evolution 1431–1442.
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.
Jennions, M.D., Petrie, M., 1997. Variation in mate choice and mating preferences: a review of causes and consequences. Biol. Rev. 72 (2), 283–327.
Kurokawa, S., 2017. The extended reciprocity: strong belief outperforms persistence.. J. Theor. Biol. 421, 16–27.
Mathew, S., Boyd, R., 2009. When does optional participation allow the evolution of cooperation? Proc. R. Soc.London B 276 (1659), 1167–1174.
McNamara, J.M., Barta, Z., Fromhage, L., Houston, A.I., 2008. The coevolution of choosiness and cooperation. Nature 451 (7175), 189–192.
Miller, R.R., 1967. No play: a means of conflict resolution.. J. Personality Social Psychol. 6 (2), 150.
Molander, P., 1985. The optimal level of generosity in a selfish, uncertain environment. J. Conflict Resolut. 29 (4), 611–618.
Noë, R., Hammerstein, P., 1994. Biological markets: supply and demand determine the effect of partner choice in cooperation, mutualism and mating. Behav. Ecol. Sociobiol. 35 (1), 1–11.
Nowak, M., Sigmund, K., 2005. Evolution of indirect reciprocity. Nature 1291–1298.
Nowak, M.A., Sigmund, K., 1998a. The dynamics of indirect reciprocity. J. Theor. Biol. 194 (4), 561–574.
Nowak, M.A., Sigmund, K., 1998b. Evolution of indirect reciprocity by image scoring. Nature 393 (6685), 573–577.
Orbell, J.M., Dawes, R.M., 1993. Social welfare, cooperators' advantage, and the option of not playing the game. Am.Sociological Rev. 787–800.
Panchanathan, K., Boyd, R., 2003. A tale of two defectors: the importance of standing for evolution of indirect reciprocity. J. Theor. Biol. 224 (1), 115–126.
Reed, L.I., Zeglen, K.N., Schmidt, K.L., 2012. Facial expressions as honest signals of cooperative intent in a one-shot anonymous prisoner's dilemma game. Evol. Human Behav. 33 (3), 200–209.
Sherratt, T.N., Roberts, G., 1998. The evolution of generosity and choosiness in cooperative exchanges. J. Theor. Biol. 193 (1), 167–177.
Sigmund, K., 2010. The Calculus of Selfishness. Princeton University Press.
Spichtig, M., Sabelis, M.W., Egas, M., 2013. Why conditional cooperators should play prisoner?s dilemma games instead of opting out. In: Evolution of Altruism: Exploring Adaptive Landscapes, p. 37.
Stanley, E.A., Ashlock, D., Smucker, M.D., 1995. Iterated prisoner's dilemma with choice and refusal of partners: evolutionary results. In: European Conference on Artificial Life. Springer, pp. 490–502.
Suzuki, Y., Toquenaga, Y., 2005. Effects of information and group structure on evolution of altruism: analysis of two-score model by covariance and contextual analyses. J. Theor. Biol. 232 (2), 191–201.
Trivers, R.L., 1971. The evolution of reciprocal altruism. Q. Rev. Biol. 46 (1), 35–57.
Vanberg, V.J., Congleton, R.D., 1992. Rationality, morality, and exit.. Am. Political Sci. Rev. 86 (02), 418–431.
Verrell, P.A., 1998. The evolution of mating systems in insects and arachnids. Am. Zool. 38 (3), 585–587.
Weibull, J.W., 1995. Evolutionary Game Theory. 1995. Massachusetts Institute of Technology.
Wubs, M., Bshary, R., Lehmann, L., 2016. Coevolution between positive reciprocity, punishment, and partner switching in repeated interactions. Proc. R. Soc. London B. 283 (1832) 20160488.
Yamagishi, T., Kikuchi, M., Kosugi, M., 1999. Trust, gullibility, and social intelligence. Asian J. Social Psychol. 2 (1), 145–161.
Zahavi, A., 1975. Mate selection?a selection for a handicap. J. Theor. Biol. 53 (1), 205–214.
Zheng, X.-D., Li, C., Yu, J.-R., Wang, S.-C., Fan, S.-J., Zhang, B.-Y., Tao, Y., 2017. A simple rule of direct reciprocity leads to the stable coexistence of cooperation and defection in the prisoner's dilemma game. J. Theor. Biol. 420, 12–17. http://dx.doi.org/10.1016/j.jtbi.2017.02.036.