



By-product mutualism with evolving common enemies

Kris De Jaegher

Utrecht University School of Economics, Utrecht University, Utrecht, The Netherlands



ARTICLE INFO

Keywords:

Evolutionary game theory
By-product mutualism
Harsh environments
Common enemies

ABSTRACT

The common-enemy hypothesis of by-product mutualism states that organisms cooperate when it is in their individual interests to do so, with benefits for other organisms arising as a by-product; in particular, such cooperation is hypothesized to arise when organisms face the common enemy of a sufficiently adverse environment. In an evolutionary game where two defenders can cooperate to defend a common resource, this paper analyzes the common-enemy hypothesis when adversity is endogenous, in that an attacker sets the number of attacks. As a benchmark, we first consider exogenous adversity, where adversity is not subject to evolution. In this case, the common-enemy hypothesis is predicted when the degree of complementarity between defenders' defensive efforts is sufficiently low. When the degree of complementarity is high, the hypothesis is predicted only when cooperation costs are high; when cooperation costs are instead low, a competing hypothesis is predicted, where adversity discourages cooperation. Second, we consider the case of endogenous adversity. In this case, we continue to predict the competing hypothesis for a high degree of complementarity and low cooperation costs. The common-enemy hypothesis, however, only continues to be predicted for the lowest degrees of complementarity.

1. Introduction

Among several explanations for cooperation among organisms (for overviews, see Dugatkin, 1997, 2002a; Sachs et al., 2004; Nowak, 2006), by-product mutualism (West Eberhard, 1975; Brown, 1983) provides a particularly straightforward rationale: organisms cooperate when it is in their individual interests to do so, and the benefits that cooperation generates for other organisms merely arise as a by-product. The common-enemy hypothesis of by-product mutualism argues that by-product mutualism particularly applies when organisms face the “common enemy of a sufficiently adverse environment” (Mesterton-Gibbons and Dugatkin, 1992, p.273), where the literature gives diverse examples of adverse environments. Increased predation risk could induce prey to jointly defend against predators (Mesterton-Gibbons and Dugatkin, 1992, p.274; Spieler, 2003; Krams et al., 2010). Predators may engage in collective hunting when facing the adverse environment of a large and difficult-to-catch prey (Scheel and Packer, 1991; Mesterton-Gibbons and Dugatkin, 1992; Dugatkin, 2002b). Further suggested examples of adverse environments that induce cooperation include scarcity in the availability of resources (Strassman et al., 2000; Callaway et al., 2002), and harsh weather conditions (Dugatkin, 1997, p. 84). Finally Roberts (2005) links adverse environments to a higher degree of interdependence between cooperating organisms.

In examples where the common enemy takes the form of the

physical environment, such as bad weather conditions, the level of adversity is exogenously given, in that it does not itself respond to the level of cooperation among the cooperating organisms (*exogenous adversity*). Yet, when the level of adversity is determined by the behavior or characteristics of another organism, such as the intensity with which a predator hunts in case of cooperatively defending prey, or the size of a prey in case of cooperatively hunting predators, the level of adversity may itself be subject to evolution, and may adapt to the level of cooperation among the cooperating organisms (*endogenous adversity*). Our paper shares the purpose of making adversity endogenous with Arenas et al. (2011). These authors extend the standard multi-player public goods game, by introducing a third strategy in the form of a “joker strategy”, on top of the standard strategies of cooperating and of defecting. Jokers are assumed to always have the same payoff, and reduce the value of the public good by a fixed amount. Whereas in the absence of jokers only joint defection can evolve, the presence of jokers can lead to rock-scissors-papers dynamics, where a fraction of the population cooperates at any given point of time.

Our analysis differs from Arenas et al. (2011), in that we instead turn a variant of the standard public goods game into an asymmetric game, by adding a population of adversaries who are matched to the population of players playing the public goods game, who are worse off the higher the value of the public good produced by the players to whom they are matched, and who can either make few or many attempts to reduce the value of this public good. This alternative

E-mail address: k.dejaegher@uu.nl.

<http://dx.doi.org/10.1016/j.jtbi.2017.02.029>

Received 14 September 2016; Received in revised form 3 February 2017; Accepted 23 February 2017

Available online 01 March 2017

0022-5193/ © 2017 Elsevier Ltd. All rights reserved.

approach allows us to investigate the following question: if increased adversity, through the common-enemy effect, makes a first group of organisms (e.g. prey) cooperate more often, should a second group of organisms (e.g., predators) that determines the level of adversity and that has lower fitness the higher the level of cooperation, then not evolve to keep adversity limited, thus preventing the common-enemy effect from coming into force?

We investigate this question by adapting the game-theoretic model of collective defense by De Jaegher and Hoyer (2016a). In this model, two defenders face a number of random attacks, and individually decide either to cooperate (= defend) or to defect (= not defend). The authors investigate the effect of an increase in the number of attacks. Results depend on the degree of complementarity between defenders' defensive efforts, i.e. the degree to which each defender's defensive effort is critical in ensuring collective defense. When the degree of complementarity is high, for high cooperation costs, the size of the basin of attraction of an evolutionary stable strategy (or ESS; Maynard Smith and Price, 1973) where both defenders cooperate is larger the higher the number of attacks (common-enemy effect). This is because the dominant effect of an increase in the number of attacks is that it becomes less attractive for defenders to deviate from joint cooperation. When the degree of complementarity is high but cooperation costs are instead low, the size of the basin of attraction of an ESS where both defenders cooperate is smaller the higher the number of attacks (competing effect). This time, the dominant effect of an increase in the number of attacks is that it becomes less attractive for defenders to deviate from joint defection. For lower degrees of complementarity, a higher number of attacks makes it more attractive to defend independently of the other defender's behavior, and the common-enemy effect is always obtained.

As shown in the current paper, when the number of attacks is endogenous in being set by an attacker, the competing effect continues to be predicted for high complementarity and low cooperation costs. The common-enemy effect, however, is only fully maintained for the lowest degrees of complementarity. Intuitively, let attacking costs initially be high, so that attackers attack few times. If cooperation costs are high, defenders will still always defect. If attacking costs now decrease, leading attackers to attack more often, by the reasoning above, when complementarity is high it becomes less attractive for defenders to deviate from joint cooperation, and the common-enemy effect may apply, in that joint cooperation is achieved. Yet, once defenders have achieved joint cooperation, it is no longer worthwhile for attackers to attack more often, and the common-enemy effect is undone. The common-enemy effect becomes self-defeating, in the sense that once a high number of attacks have lead to joint cooperation, attackers no longer have an incentive to launch many attacks. When complementarity is low, an increase in the number of attacks after a decrease in attacking costs, does not lead defenders to achieve joint cooperation, but only makes a higher fraction of defenders cooperate. For this reason, attackers continue to attack when the common-enemy effect applies, and the common-enemy effect is no longer self-defeating.

The paper is structured as follows. Section 2 presents the model. As a benchmark, Section 3 shortly treats the case of exogenous adversity. Section 4 contains the central results of this paper. We end with a discussion in Section 5.

2. The model

We consider the following evolutionary game played by two infinitely large populations of defenders and attackers (for convenience, an attacker is referred to as "she" and a defender as "he"). At each point of time, two defenders are randomly matched to each other, and at the same time to one randomly-chosen attacker. Each attack by an attacker is targeted at a single, randomly chosen defender among the two defenders to whom she is matched. An attacker can launch either one ($A = 1$) or two ($A = 2$) random attacks on the two defenders to

whom she is matched. When $A = 2$, the two attacks (interpreted as a process of statistical sampling of the two defenders) take place in a process of sampling with replacement, so that by coincidence the same defender may be attacked twice.

Any two defenders who are matched to each other hold a common resource, from which they always obtain the same fitness. Each defender either plays C (cooperates) or plays D (defects). Playing C means exerting effort to defend the common resource, and comes at a cost c ; playing D means not exerting any effort, and incurring zero costs. If only one defender is attacked (which occurs either when $A = 1$, or when $A = 2$ but the same defender is by chance attacked twice), then what the other defender plays does not matter for the fitness both defenders obtain from their common resource. If the solely-attacked defender plays C , both defenders obtain the maximal fitness V from the common resource; if the solely-attacked defender plays D , both defenders obtain fitness $(1 - k)V$ net of cooperation costs, with $0 < k \leq 1$, where k is the *degree of complementarity* (see below). If both defenders are attacked (which occurs when $A = 2$, and by chance a different attacker is each time attacked), when they both play C , both obtain maximal fitness V from the common resource; when one plays C and the other D , they again both obtain fitness $(1 - k)V$ from the common resource; when they both play D , they both obtain zero fitness.

Restating the model, each defender may be seen as either contributing to the preservation of the fitness V obtained from the common resource, or not contributing. A defender may contribute in two ways: either by being attacked but playing C , or by not being attacked (in which case it does not matter whether he plays C or D). A defender does not contribute when playing D and being attacked. With this restatement of the model, the parameter k is more easily interpreted as the degree of complementarity (Ray et al., 2007) between the defenders' contributions, where $k = 1$ means perfect complementarity (= only when both defenders contribute can nonzero fitness be obtained); $k = 1/2$ means that the defenders' contributions are perfect substitutes (= a second contributing defender adds as much to fitness as a first contributing one); k approaching zero means that the defenders' contributions have a best-shot nature (Hirshleifer, 1983) (= the first contributing defender produces almost the maximal fitness V , and a second contributing defender adds little).

Any attacker obtains the negative of the fitness that the two defenders to whom she is matched obtain from their common resource. Each attack comes at a cost g to an attacker. Let q denote the fraction of cooperating defenders in the defender population, and let a denote the fraction of attackers in the attacking population who attack twice. Denote by f_C a defender's average fitness from cooperating and by f_D a defender's average fitness from defecting, so that $(f_C - f_D)$ is the increase in average defender fitness from cooperating rather than defecting. Denote by $f_{A=1}$ and $f_{A=2}$ an attacker's average fitness from respectively attacking once, and attacking twice, so that $(f_{A=2} - f_{A=1})$ is the increase in average attacker fitness from attacking twice rather than once. Following the replicator dynamics (Taylor and Jonker, 1978), changes in the defender population are given by the relative performance of cooperating and of defecting, and changes in the attacker population are given by the relative performance of attacking twice and once. The evolutionary dynamics are now fully described by the following system of two differential equations:

$$\dot{q} = q(1-q)(f_C - f_D) \quad (1)$$

$$\dot{a} = a(1-a)(f_{A=2} - f_{A=1}) \quad (2)$$

While our model is stripped of complexities to an extent that it is best seen as a metaphor, it best fits territorial defense against intruders (e.g., Grinnell et al., 1995; Port et al., 2011; Gese, 2001; Rubenstein and Nuñez, 2009), or collective defense of prey against predators (e.g., Spieler, 2003). As a hypothetical example, consider two male lions defending a common territory. Let each lion guard one side of the

territory, where a cooperating lion engages any intruder on his side of the territory, whereas a defecting lion does not. An intruding lion randomly chooses one side of the territory to attack, and may perform several such random attacks. Note that when one lion's side is never attacked, it does not matter whether this lion does or does not invest in engaging. In one extreme case, when a single lion is attacked and this lion does not engage, this means that the entire territory is lost (degree of complementarity $k = 1$). Alternatively, when a lion is attacked and does not engage, non-attacked lions or attacked lions that do engage, can still partially stop the intruder, and only part of the territory is lost ($k < 1$). As one approaches the other extreme (k approaching zero), only one non-attacked lion, or attacked lion that engages, suffices to preserve almost the entire territory.

Looking at the case of perfect complementarity ($k = 1$), we note that this case bears resemblance to so-called *cycloaexy* (Jolivet et al., 1990), or circular defense, where a group of animals position themselves in a circle, with the purpose of collective defense of each other or of offspring, against a predator. The expression “circling the wagons” as a synonym for defending against a common threat, reflects the underlying mechanism: if one player breaks ranks, all players suffer.

While the game we model is framed as modeling collective defense, we note that by relabeling the strategies, our game may alternatively be seen as modeling cooperative hunting. The two symmetric players in our game are then considered as predators, who may either cooperate by taking up position in the encirclement of a prey (i.e., the third player), or defect by not taking up position (for an example of encirclement by predators, see Stander, 1992). In particular, assume that the two predators have coordinated so that one predator takes up position on the left of the prey, and the other on the right. The prey tries to escape the encirclement either once ($A = 1$) or twice ($A = 2$), each time randomly choosing to escape either on the left, or on the right. For $k = 1$, if one predator has not taken up position on the side from which the prey attempts to escape, the prey's escape is successful.

Several of our modeling assumptions deserve attention. *First*, we assume that the attacker chooses only from two attacking levels, and in particular chooses between attacking once or twice. While De Jaegher and Hoyer (2016a) in a model with exogenously given numbers of attacks consider any positive attacking level, in a model with endogenous attacks, considering many attacking levels becomes analytically intractable. The assumption that the attacker can choose between two attacking levels models in a basic manner that, in response to the level of cooperation among defenders, the attacker may choose to increase or to decrease the number of attacks. The model focuses specifically on either one or two attacks, because with one attack there is always a dominant strategy for attackers, making the effect on the probability of cooperation of an increase in the number of attacks immediately clear. As shown in De Jaegher and Hoyer (2016a), when the choice is instead between one level of multiple attacks and another such level, the change in the level of cooperation is calculated by the change in the size of the basin of attraction of an ESS where all defenders cooperate, or by change in the fraction of cooperating defenders in a mixed ESS; this makes calculating the effect of an increase in the number of attacks more difficult, a complication that is avoid in the present paper.

When limiting oneself to two attacking levels, at first sight, an even simpler model would seem to suffice, where attackers choose between attacking once and not attacking. Yet, it is clear that in the absence of attacks defenders never cooperate, whereas in the presence of attacks they might, so that the effect on the level of cooperation of an increase in the number of attacks from zero to one is always positive. Therefore, such a simplified model cannot reflect the intuition that an increase in the number of attacks may on the one hand make it less attractive to deviate from joint cooperation, but on the other hand may make it less attractive to deviate from joint defection as well, making the effect of an increase in the number of attacks ambiguous. Moreover, assuming that attackers never reduce their number of attacks to zero is realistic because, contrary to what is the case in our simplified model, joint

cooperation by a pair of defenders may sometimes fail, making it optimal for attackers to always maintain a positive number of attacks, as attacks may then be successful even if all defenders cooperate.

Second, attacks are random in our model, meaning that when one defender cooperates and the other defender defects, the attacker may still by chance attack the cooperating defender, making her attack unproductive. Alternatively, the attacker may have the cognitive ability to identify the intentions of the defenders in the group it faces, and in this manner be able to attack any defecting defenders in the group, and to avoid attacking cooperating defenders (for the role of intention recognition among cooperating players, see, e.g., Han et al., 2011). Yet, such cognitive abilities may themselves come at fitness costs to attackers. In a similar context to the present model, De Jaegher and Hoyer (2016b) study the effect of an exogenous increase in the level of adversity, measured as a switch from a random attack to a targeted attack, and obtain results that are analogous to those for an increase in the number of random attacks. Intuitively, the effect of a higher number of random attacks is analogous to the effect of higher ability to perform targeted attacks: in each case, any defecting defender is attacked with higher probability. Because of this, a model where adversity consists of a higher ability to perform targeted attacks, and where the level of adversity is endogenous, will lead to similar results as those in the current paper.

Third, our model makes the simplifying assumption that it does not matter for the value of the public good how often an individual non-defending player is attacked, but that it does matter how many different non-defending players are attacked at least once. Yet, it is realistic that, all else equal, the value of the public good is also reduced the more often an individual attacker is successfully attacked. An alternative model may now be considered where it only matters for the value of the common resource how many of the attacks were successful in targeting a defecting defender, but not how many different defenders were hit by a successful attack. It is straightforward to check that in such an alternative model, exogenous increases in adversity lead to effects analogous to those we describe below in Section 3, implying that the results for endogenous adversity are also similar in such an alternative model. This suggests that our simplifying assumption that it does not matter how often an individual defecting defender is successfully attacked, is innocuous.

Fourth, in our model, each individual attack is directed at a single defender, and the effect on the common resource of this particular attack depends only on the play of this defender, and not on the play of other defenders. In an alternative model, one could consider a group of defenders as being collectively attacked, where the probability of successful defense of the common resource is higher the more defenders in the group contribute to collective defense. An example here may be mobbing of a predator by a group of prey (e.g. Sandoval and Wilson, 2012). Yet, the effect of a higher attacking intensity in such an alternative model may be to make the defenders' contributions to collective defense more complementary. As shown in De Jaegher and Hoyer (2016a), an increase in the degree of complementarity in the production of a common good has similar effects as a higher number of attacks in the current model.

3. Benchmark: exogenous adversity

As a benchmark, let us look at the case where the level of A is exogenously given (exogenous adversity), so that only Eq. (1) is relevant. The game studied is now a symmetric two-player public goods game. To characterize what form this public goods game takes as a function of the number of attacks, we first shortly repeat the taxonomy of symmetric two-player public goods games (cf. Doebeli and Hauert, 2005; Archetti and Scheuring, 2012), which are represented generally in Eq. (3). The reward each player receives when both players cooperate is represented as R . The temptation payoff one defecting player obtains when the other player cooperates is denoted T .

Table 1

Four public-goods games, with payoffs R (reward), S (sucker), T (temptation) and P (punishment). It is assumed that both R and T are larger than S and P . Different games are obtained as a function of the sign of $(R - T)$ (respectively $(S - P)$), defined as the increase in fitness of cooperating rather than defecting when the other player cooperates (respectively, defects).

		$(S - P)$	
		< 0	> 0
$(R - T)$	< 0	Prisoner's Dilemma	Snowdrift game
	> 0	Stag Hunt	Harmony Game

The sucker payoff one cooperating player obtains when the other player defects is denoted S . Finally, the punishment payoff each player receives when both players defect is denoted P . It is assumed that both R and T are larger than S and P . This means that the individual player is always better off if the other player cooperates rather than defects, that in a pair with a single cooperating player it is better to be the defecting player, and that joint cooperation leaves players better off than joint defection.

$$\begin{matrix} & C & D \\ C & \begin{pmatrix} R & S \\ T & P \end{pmatrix} \\ D & \end{matrix} \quad (3)$$

The type of the two-player public goods game is determined by the sign of $(R - T)$ (increase in fitness of cooperating rather than defecting when the other player cooperates), and the sign of $(S - P)$ (increase in fitness of cooperating rather than defecting when the other player defects), leading to four types of games as represented in Table 1. The game is a Prisoner's Dilemma (Tucker, 1950) when $(R - T) < 0$ and $(S - P) < 0$. In this case, the unique ESS is joint defection. The game is a Harmony Game when $(R - T) > 0$ and $(S - P) > 0$. In this case, the unique ESS is joint cooperation.

The game is a Stag Hunt (Skyrms, 2004) when $(R - T) > 0$ and $(S - P) < 0$ (which is only possible if the increase in fitness of cooperating rather than defecting is larger when the other player cooperates than when he defects). In this case, both joint cooperation and joint defection are ESS's (additionally, there is a mixed Nash equilibrium which is not an ESS). In a Stag Hunt, it is in the interest of the individual player to cooperate when a sufficient fraction of the population cooperates, and to defect when a sufficient fraction of the population also defects. As both ESS's have positive basins of attraction, we assume that they are both played with positive probability (assuming that each initial population is equally likely, the probability that one of the ESS's is played is equal to the size of its basin of attraction (De Jaegher and Hoyer, 2016a)).

Finally, the game is a Snowdrift game (Sugden (1986); also called Hawk-Dove game (Maynard Smith and Price, 1973), or Chicken game (Russell, 1959)) when $(R - T) < 0$ and $(S - P) > 0$ (which is only possible if the increase in fitness of cooperating rather than defecting is larger when the other player defects than when he cooperates). In this case, a mixed equilibrium is the unique ESS, where a specific fraction of the population cooperates (specifically this fraction q must meet the condition $qR + (1 - q)S = qT + (1 - q)P$ so that $q = (S - P)/(T - R + S - P)$). In a Snowdrift game, it is in the interest of the individual player to defect when a sufficient fraction of the population cooperates, and it is in the interest of the individual player to cooperate when a sufficient fraction of the population defects.

We now turn to the specific defender game in our model. Eq. (4) represents the defender game with $A = 1$, and (5) the defender game with $A = 2$, where each time rows represent the actions of a focal defender, and columns the actions of the other defender. The payoffs in these games are explained as follows. Let us first look at the case where one defender cooperates and the other defects. If the number of attacks now equals one, with probability $1/2$ only the cooperating defender is

attacked, and both defenders obtain fitness V from the common resource. If the number of attacks is fixed at two, the probability that only the cooperating defender is attacked equals $1/2 * 1/2 = 1/4$. It follows that the complementary probability that the defecting defender is attacked at least once, and that both defenders obtain fitness $(1 - k)V$ from the resource, equals $1/2$ for one attack, and $3/4$ for two attacks. We conclude that when one defender cooperates and one defects, average fitness net of cooperation costs equals $1/2V + 1/2(1 - k)V = (1 - 1/2k)V$ with one attack, and $1/4V + 3/4(1 - k)V = (1 - 3/4k)V$ in case of two attacks.

Next, let us look at the case where both defenders defect. If the number of attacks now equals one, with probability one a single defecting defender is attacked, meaning that both defenders obtain expected fitness $(1 - k)V$ from the common resource. If the number of attacks equals two, one specific defender is attacked twice with probability $1/2 * 1/2 = 1/4$; it follows that the event of only one defender being attacked occurs with probability $1/4 + 1/4 = 1/2$. In this case, both defenders obtain fitness $(1 - k)V$ from the common resource. With the complementary probability, both defenders are attacked, in which case both obtain payoff 0. It follows that the expected fitness net of cooperation costs obtained by both defenders when facing two attacks, equals $1/2(1 - k)V$. Finally, if both defenders cooperate, they always each obtain fitness V from the common resource.

$$\begin{matrix} & C & D \\ C & \begin{pmatrix} V - c & (1 - 1/2k)V - c \\ (1 - 1/2k)V & (1 - k)V \end{pmatrix} \\ D & \end{matrix} \quad (4)$$

$$\begin{matrix} & C & D \\ C & \begin{pmatrix} V - c & (1 - 3/4k)V - c \\ (1 - 3/4k)V & 1/2(1 - k)V \end{pmatrix} \\ D & \end{matrix} \quad (5)$$

In the manner of Mesterton-Gibbons and Dugatkin (1992, 1997), we want to know how an increase in the number of attacks affects the evolution of defender play: the switch from $A = 1$ and $A = 2$ may then indeed create the “common enemy of a sufficiently harsh environment”, inducing defenders to cooperate more often. To see whether such a common-enemy effect exists in the model, we need to know how the ESS's are affected by the number of attacks. As shown in De Jaegher and Hoyer (2016a), depending on the number of attacks, but also on the degree of complementarity k and the cooperation costs c , the defender game takes on different forms, distinguished by the types and number of ESS's (see the taxonomy in Table 1).

In the defender game with one attack (see (4)), it is the case that $(R - T) = (S - P) = 1/2kV - c$. Because $(R - T)$ and $(S - P)$ always have the same sign, by Table 1, with one attack the defender game can only either be a Harmony game ($c < 1/2kV$), or a Prisoner's Dilemma ($c > 1/2kV$). In the defender game with two attacks (see (5)), $(R - T)$ typically differs from $(S - P)$. In particular, by (5), $(R - T) = 3/4kV - c$, and $(S - P) = (1/2 - 1/4k)V - c$. It follows that for sufficiently high complementarity ($k > 1/2$), it is the case that $(R - T) > (S - P)$, and cost ranges exist such that the defender game is a Stag Hunt. For sufficiently low complementarity ($k < 1/2$), it is the case that $(R - T) < (S - P)$, and cost ranges exist such that the defender game is a Snowdrift game.

Proposition 1 now analyzes the effect of an increase in the number of attacks on the type of game played, and indicates after the \Rightarrow -sign whether this implies a common-enemy effect, or instead a competing effect where higher adversity discourages cooperation rather than encouraging it.¹ Three main scenarios are distinguished (high, intermediate and low complementarity), and within each of these scenarios two cases (low and high cooperation costs), which are precisely defined

¹ In an experiment Polizzi di Sorrentino et al. (2012) investigate how between-group hostility affects cooperation in capuchin monkeys, the authors propose both a cooperative hypothesis fitting our common-enemy effect, and an induced-tension hypothesis that fits our competing effect, and find evidence for the latter hypothesis. In a similar experiment with cooperative fish, however, Bruinjes et al. (2016) find evidence for the former hypothesis.

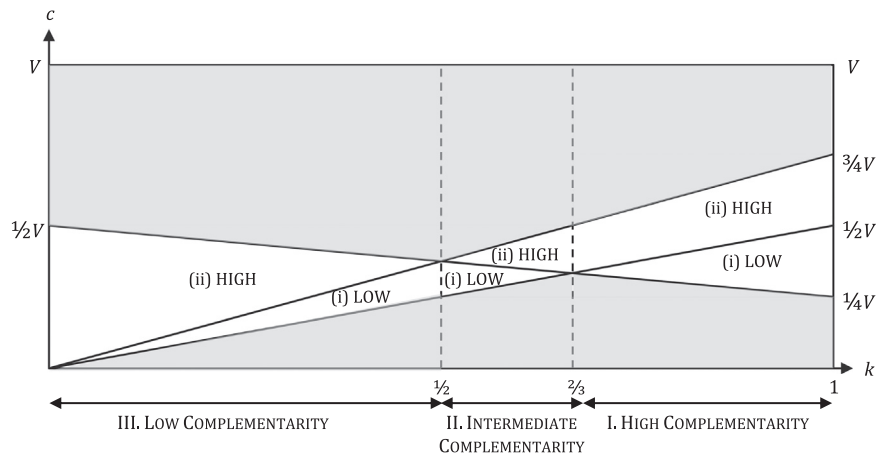


Fig. 1. Representation of the cases in Definition 1, with cooperation costs c on the Y-axis, degree of complementarity k on the X-axis. Cases (i) low cooperation costs and (ii) high cooperation costs as a function of the degree of complementarity, where we distinguish between scenarios (I) high complementarity, (II) intermediate complementarity, and (III) low complementarity.

in Definition 1. The parameter ranges considered are represented in Fig. 1, exemplifying that what constitutes low and high cooperation costs depends on the scenario. Furthermore, we ignore cooperation costs that are so high that the game remains a Prisoner's Dilemma (upper grey area in Fig. 1), or where the costs are so low that it remains a Harmony game, whatever the number of attacks (lower grey area in Fig. 1).

Definition 1. High, low and intermediate complementarity (scenarios); low and high cooperation costs (cases):

- I. For high complementarity ($\frac{2}{3} < k < 1$), define (i) $(\frac{1}{2} - \frac{1}{4}k)V < c < \frac{1}{2}kV$ as low cooperation costs, and (ii) $\frac{1}{2}kV < c < \frac{3}{4}kV$ as high cooperation costs.
- II. For intermediate complementarity ($\frac{1}{2} < k < \frac{2}{3}$), define (i) $\frac{1}{2}kV < c < (\frac{1}{2} - \frac{1}{4}k)V$ as low cooperation costs, and (ii) $(\frac{1}{2} - \frac{1}{4}k)V < c < \frac{3}{4}kV$ as high cooperation costs.
- III. For low complementarity ($0 < k < \frac{1}{2}$), define (i) $\frac{1}{2}kV < c < \frac{3}{4}kV$ as low cooperation costs, and (ii) $\frac{3}{4}kV < c < (\frac{1}{2} - \frac{1}{4}k)V$ as high cooperation costs.

The results in Proposition 1 are represented in Fig. 2 in the form of phase spaces, each time both for one attack (bottom axis) and two attacks (top axis), and this for each of the two cases (low and high cooperation costs) within each of the three scenarios (high, intermediate and low complementarity). The axes represent all possible defender population states (q ranging from zero to 1), arrows represent the direction in which the defender population evolves. Open circles represent fixed points that are not ESS's, and solid circles represent fixed points that are ESS's (note that by (1), both $q = 0$ and $q = 1$ are always fixed points). Each time it is indicated whether the common-enemy effect or the competing effect applies for a change from one to two attacks.

The results may be understood by looking at the effect of an increase in the number of attacks on $(R - T)$ and on $(S - P)$. Intuitively, for high complementarity, $(R - T)$ increases for a higher number of attacks, as the consequences of unilaterally deviating from a situation where all defenders cooperate are then more severe; at the same time, $(S - P)$ decreases for a higher number of attacks, as the benefit of unilaterally cooperating in a situation where all defenders defect is then smaller. This is why the game becomes a Stag Hunt when the number of attacks is increased. As with only one attack $(R - T) = (S - P)$, the effect of switching to two attacks depends on whether cooperation costs are low (Fig. 2.I(i), so that the starting point is a Harmony game), or high (Fig. 2.I(ii), so that the starting point is a Prisoner's Dilemma). The common-enemy effect and competing effect apply here respectively for high and low cooperation costs, based on the premise that in the

Stag Hunt that is obtained for two attacks, both ESS's are achieved with positive probability (as seems plausible when they each have a sizeable basin of attraction).

For intermediate complementarity, it continues to be true that $(R - T) > (S - P)$ for two attacks (meaning that in any pair of defenders, a second contributing defender adds more to the common resource than a first contributing defender), but these expressions lie closer to each other, and both increase compared to the case of one attack. For this reason, as the number of attacks is increased, the game either switches from a Prisoner's Dilemma to a Harmony game (Fig. 2.II(i)), or from a Prisoner's Dilemma to a Stag Hunt (Fig. 2.II(ii)). For low complementarity, it is instead the case that $(R - T) < (S - P)$ for two attacks (meaning that in any pair of defenders, a second contributing defender adds less to the common resource than a first contributing defender), where these expressions again both increase compared to the case of one attack. It follows that, for an increase in the number of attacks, the game either switches from a Prisoner's Dilemma to a Harmony game (Fig. 2.III(i)), or from a Prisoner's Dilemma to a Snowdrift game (Fig. 2.III(ii)).

Proposition 1. Exogenous adversity: defender games played as a function of exogenously given number of attacks, and implications for the relevance of the common-enemy effect/competing effect.

- I. High complementarity:
 - (i) Low cooperation costs: (a) Harmony game for one attack; (b) Stag Hunt for two attacks. \Rightarrow competing effect.
 - (ii) High cooperation costs: (a) Prisoner's Dilemma for one attack; (b) Stag Hunt for two attacks. \Rightarrow common-enemy effect.
- II. Intermediate complementarity:
 - (i) Low cooperation costs: (a) Prisoner's Dilemma for one attack; (b) Harmony game for two attacks. \Rightarrow common-enemy effect.
 - (ii) High cooperation costs: (a) Prisoner's Dilemma for one attack; (b) Stag Hunt for two attacks. \Rightarrow common-enemy effect.
- III. Low complementarity:
 - (i) Low cooperation costs: (a) Prisoner's Dilemma for one attack; (b) Harmony game for two attacks. \Rightarrow common-enemy effect.
 - (ii) High cooperation costs: (a) Prisoner's Dilemma for one attack; (b) Snowdrift game for two attacks. \Rightarrow common-enemy effect.

Proof: see De Jaegher and Hoyer (2016a).

4. Endogenous adversity: results

We now analyze the asymmetric attacker-defender game originally set out in Section 2, where adversity in the form of the number of

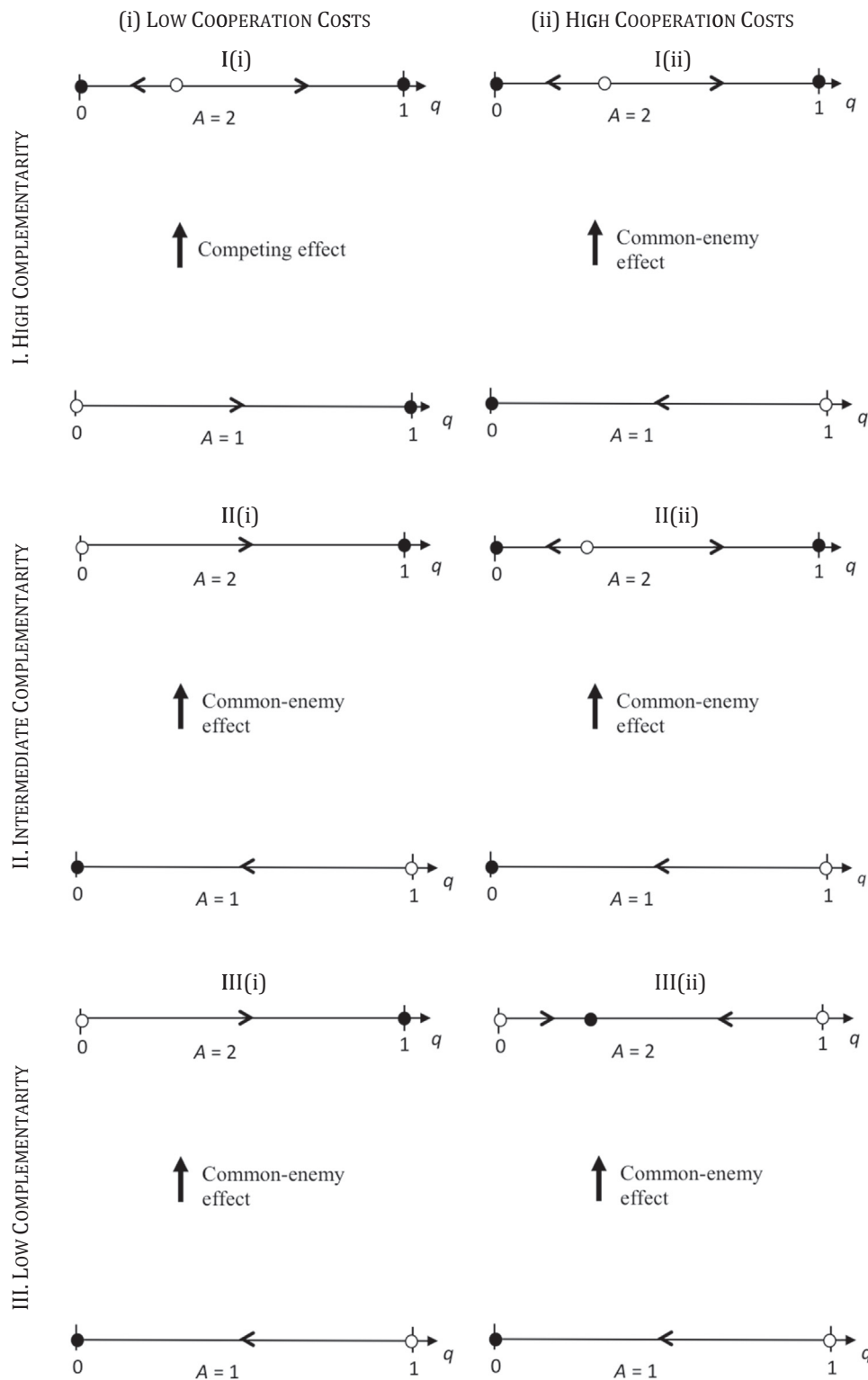


Fig. 2. Exogenous adversity. For six combinations of three scenarios (high, intermediate and low complementarity) and two cases (low and high cooperation costs), the phase space is represented both for one (bottom) and two attacks (top). Arrows on the axes indicate the direction in which the defender population evolves; solid circles denote ESS's, and open circles fixed points that are not ESS's. For each case, it is specified whether the common-enemy effect applies, or the competing effect.

random attacks arises endogenously, as it is determined by an attacker population. Our analysis builds up towards [Proposition 2](#), which has exactly the same structure as [Proposition 1](#). The difference with the model of exogenous adversity in [Section 3](#) is that, instead of looking for each of the cases at the effect of an exogenous increase in the number of attacks (from case (a) with one attack to case (b) with two attacks), and drawing conclusions on whether the common-enemy effect or the

competing effect applies, in the model with endogenous adversity we look instead at the effect of a decrease in attackers' attacking costs (from case (a) with high attacking costs to case (b) with low attacking costs), to check whether similar conclusions continue to apply. Intuitively, on the one hand it is conceivable that the model with endogenous adversity operates in the same manner as the model with exogenous adversity: in this reasoning, a decrease in the attacking costs

inexorably leads attackers to attack more often, and following Proposition 1 leads to either a competing effect, or a common-enemy effect. Yet, on the other hand, it is also conceivable that the model with endogenous adversity works differently: specifically, if a larger number of attacks causes a common-enemy effect, attackers could then evolve to keep the number of attacks low even when attacking costs are low, to avoid the common-enemy effect from operating.

To see which of these intuitions is correct, as attackers evolve to set the level of attacking that maximizes their fitnesses, for each of the two cases within the three scenarios, we need to derive the asymptotically stable states² both when attacking costs are high, and when they are low. As a tool for a dynamic analysis of the attacker-defender game, one can represent all possible population states in a phase space with q on the X-axis and a on the Y-axis. We first characterize fixed points of the dynamic system (1) and (2), where neither q nor a changes, and then consider the asymptotic stability of these fixed points. In order to find the fixed points, we characterize the nullclines of (1) and (2). A defender (respectively attacker) nullcline is a curve in the phase space such that q (respectively a) does not change, i.e. such that the right-hand side of (1) (respectively (2)) is equal to zero. Graphically, we are then able to characterize all fixed points of (1) and (2) by finding all the points where a defender nullcline and an attacker nullcline intersect. Moreover, such nullclines allow us to derive for every population state in the phase space in what directions the populations evolve, as this depends on the position of the population state relative to the nullclines.

Trivial nullclines are the edges of the phase space, i.e. the defender nullclines $q = 0$ and $q = 1$, and the attacker nullclines $a = 0$ and $a = 1$. In Lemmata 1 and 2, we derive the position of the non-trivial nullclines (i.e. other than the edges of the phase space). We start in Lemma 1 by deriving the position in the phase space of the non-trivial defender nullcline, where $(f_C - f_D) = 0$. Several cases can be distinguished, as represented in Fig. 3.

The results in Lemma 1 can easily be explained by means of the results for exogenous adversity in Proposition 1, and by means of the corresponding Fig. 2. For each individual case, the phase space for $A = 1$ in Fig. 2, should be identical to the part of the phase space in Fig. 3 where $a = 0$; in the same way, the phase space for $A = 2$ in Fig. 2, should be identical to the part of the phase space in Fig. 3 where $a = 1$. Thus, Fig. 3 can be seen as an extension of Fig. 2, representing not only the cases $a = 0$ and $a = 1$, but any level of a . As a defender non-trivial nullcline is a curve dividing the areas in the phase space where the fraction of cooperating defenders in the population decreases and increases, the position of the nullclines in Fig. 3 follows directly from Fig. 2, where Appendix A additionally shows that all defender nullclines are monotonous.

Lemma 1. Endogenous adversity: position in phase space of non-trivial defender nullcline.

I. High complementarity:

- (i) Low cooperation costs: nullcline increasing, cuts the edges $q = 0$, $a = 1$.
- (ii) High cooperation costs: nullcline decreasing, cuts the edges $q = 1$, $a = 1$.

II. Intermediate complementarity:

- (i) Low cooperation costs: nullcline decreasing, cuts the edges $q = 0$, $q = 1$.
- (ii) High cooperation costs: nullcline decreasing, cuts the edges $q = 1$, $a = 1$.

² In Section 2, we use the more straightforward ESS concept, as for symmetric games, the set of ESS's coincides with the set of asymptotically stable states. As in asymmetric games, there may be asymptotically stable states that are not ESS, the stability concept in Section 3 is asymptotic stability.

III. Low complementarity:

- (i) Low cooperation costs: nullcline increasing, cuts the edges $q = 0$, $q = 1$.
- (ii) High cooperation costs: nullcline increasing, cuts the edges $q = 0$, $a = 1$.

Proof: see Appendix A.

Next, in Lemma 2, we investigate the position in the phase space of non-trivial attacker nullclines, where $(f_{A=2} - f_{A=1}) = 0$. The cases considered are defined in Definition 2, and represented in Fig. 4. To maintain comparability between high (low) attacking costs in the model with endogenous adversity on the one hand, and a low (high) number of attacks in the model with exogenous adversity on the other hand, we distinguish between (a) high attacking costs, and (b) low attacking costs. For high complementarity, as two subcases of high attacking costs lead to qualitatively different results, we further distinguish between (a') the upper range of high attacking costs, and (a'') the lower range of high attacking costs.

Definition 2. High, low and intermediate complementarity; high and low attacking costs.

- I. For high complementarity ($\frac{2}{3} < k < 1$), define (a') $\frac{1}{2}V[\frac{1}{2}k]^2/(2k-1) < g < kV$ as the upper range of high attacking costs, (a'') $\frac{1}{2}(1-k)V < g < (\frac{1}{2}V)[\frac{1}{2}k]^2/(2k-1)$ as the lower range of high attacking costs, and (b) $0 < g < \frac{1}{2}(1-k)V$ as low attacking costs;.
- II. /III. For intermediate/low complementarity ($0 < k < \frac{2}{3}$), define (a) $g > \frac{1}{2}(1-k)V$ as high attacking costs, and (b) $0 < g < \frac{1}{2}(1-k)V$ as low attacking costs.

Each case in Lemma 2 is represented in a corresponding part of Fig. 5. Intuitively, if all defenders cooperate, then attacks are unproductive so that attacking once is always better than attacking twice. For this reason, as long as attacking costs are positive, there is always a range of high fractions of cooperating defenders, such that attackers are better off attacking once; indeed, in each case in Fig. 5, on the right vertical edge of the phase space, as indicated by a downward-pointing arrow, the attacker population evolves towards attacking once. For a range of high attacking costs, attacking once will in fact *always* be better (Lemma 2.I(a') and 2.II/III(a)); in this case, there is no non-trivial attacker nullcline, as illustrated in parts I(a') and II/III(a) of Fig. 5, and the attacker population evolves towards attacking once in the entire phase space. For a range of low attacking costs (Lemma 2.I(b) and 2.II/III(b)), when all defenders defect, attackers are better off attacking twice rather than once; the same is true for a fraction of cooperating defenders that is sufficiently small. Since, at the same time, attacking once is better for a sufficiently large fraction of cooperating defenders, in this case there is a single non-trivial attacker nullcline, taking the form of a vertical line. As illustrated in part II/III(b) of Fig. 5, with an attacker nullcline $q = q^*$, this means that for a range of low (high) fractions of cooperating defenders, where $q < q^*$ (respectively $q > q^*$), attacking twice (respectively once) yields highest fitness (this part of the figure represents two other attacker nullclines $q = q^{**}$ and $q = q^{***}$, for different levels of attacking costs). The same is true in part I(b) of Fig. 5, to the left and to the right of q_H'' (this part of the figure represents one other attacker nullcline q_H''' , for a different level of attacking costs).

In a separate case (Lemma 2, case I(a'')), there are two non-trivial attacker nullclines. This occurs when attacking costs are sufficiently high for attacking twice not to be better to attackers when the fraction of cooperating defenders is low, where at the same time these attacking costs are not so high that attacking twice is never better ("lower range of high attacking costs"). In particular, when the degree of complementarity is sufficiently high, with a low fraction of cooperating defenders, one attack suffices to reduce the fitness obtained from the common resource considerably. At the same time, as long as the fraction of cooperating defenders is sufficiently close to 1, attacking once will also be better to attackers. Only for intermediate fractions of cooperating defenders is attacking twice then better. The resulting two

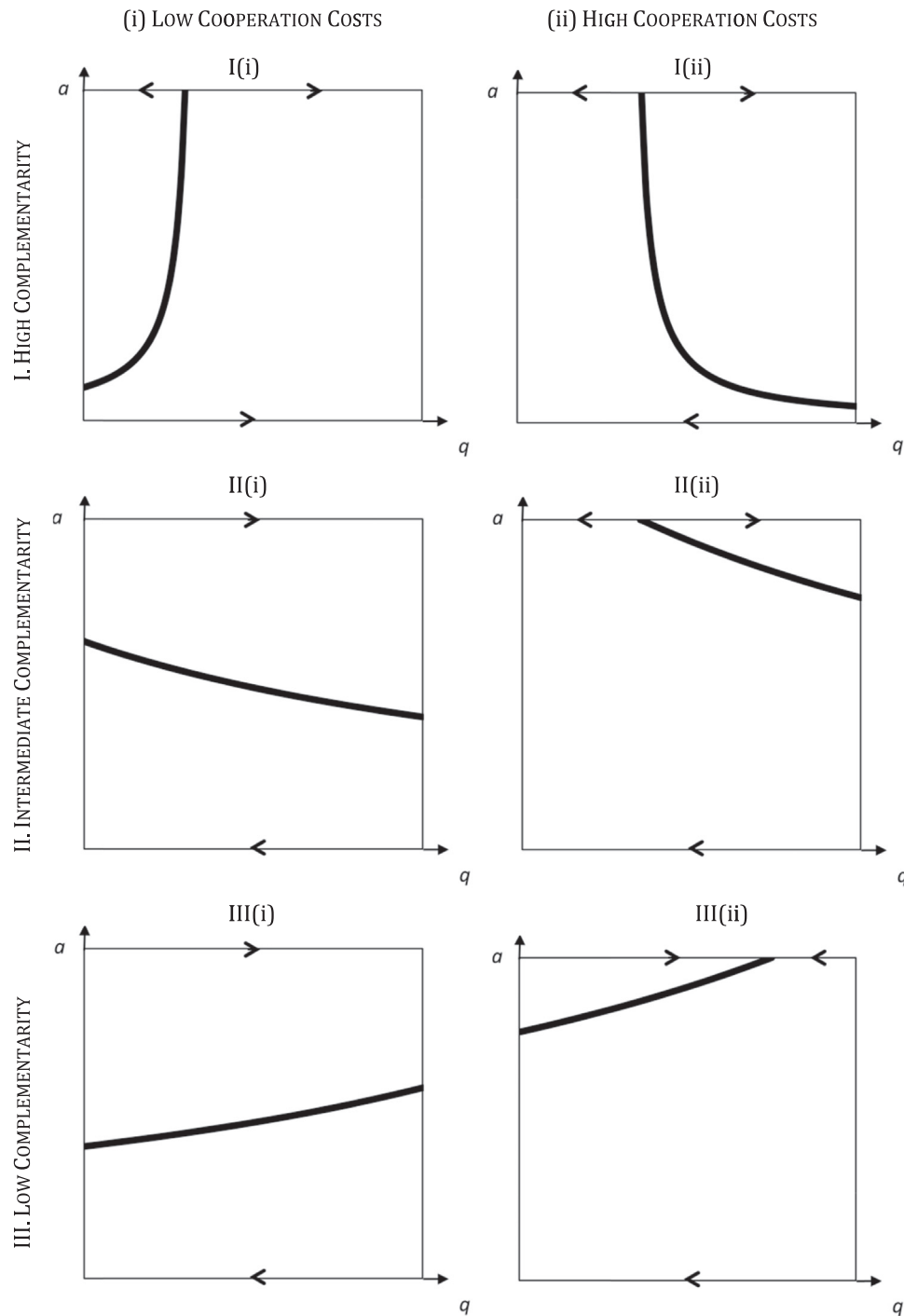


Fig. 3. Position of non-trivial defender nullclines, in the phase space with the fraction of attackers attacking twice on the Y-axis (a) and the fraction of cooperating defenders on the X-axis (q); arrows indicate direction in which defender population evolves, where the nullcline separates the area of the phase space where the population evolves in one or the other direction.

non-trivial attacker nullclines are represented in part I(a)'' of Fig. 5, and take the form of two vertical lines $q=q_L$ and $q=q_H$, with $0 < q_L < q_H < 1$; this means that for low ($q < q_L$) and high ($q > q_H$) fractions of cooperating defenders, attacker fitness is highest when attacking once, whereas for intermediate fractions ($q_L < q < q_H$), it is highest when attacking twice.

Part I(a)'' of Fig. 5 also illustrates that the lower attacker nullcline shifts to the left ($q = q_L'$), and the higher nullcline to the right ($q = q_H'$) as attacking costs are lowered. More generally, within the case of the lower range of high attacking costs, for the highest attacking costs within this range, q_L and q_H both approach the same intermediate level

of q , and for the lowest attacking costs within this range, q_L approaches zero. As one moves to the case of low attacking costs (part I(b) of Fig. 5), only the higher non-trivial attacker nullcline remains relevant, and shifts to the right as the attacking costs are further lowered, as illustrated by the nullclines $q = q_H''$ and $q = q_H'''$.

In the same manner, one can look also at the effect of lowering attacking costs on the non-trivial attacker nullclines for the intermediate- and low-complementarity scenarios. In this case, there is no intermediate attacking cost range for which there are two such nullclines. Starting from the case of high attacking costs (part II/III(a) of Fig. 5), when we decrease attacking costs to enter the case of

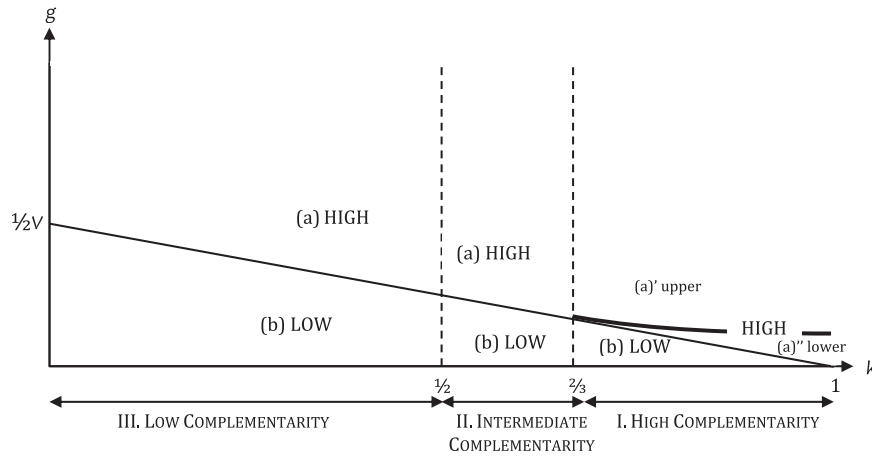


Fig. 4. Representation of the cases in Definition 2, with attacking costs g on the Y-axis, and degree of complementarity k on the X-axis. (a) High attacking costs (above straight line) and (b) low attacking costs (below straight line) as function of degree of complementarity, where we distinguish between the scenarios (I) high, (II) intermediate and (III) low complementarity. For high complementarity we further distinguish between (a)' an upper range of high attacking costs (above curve), and (a)'' a lower range of high attacking costs (below curve).

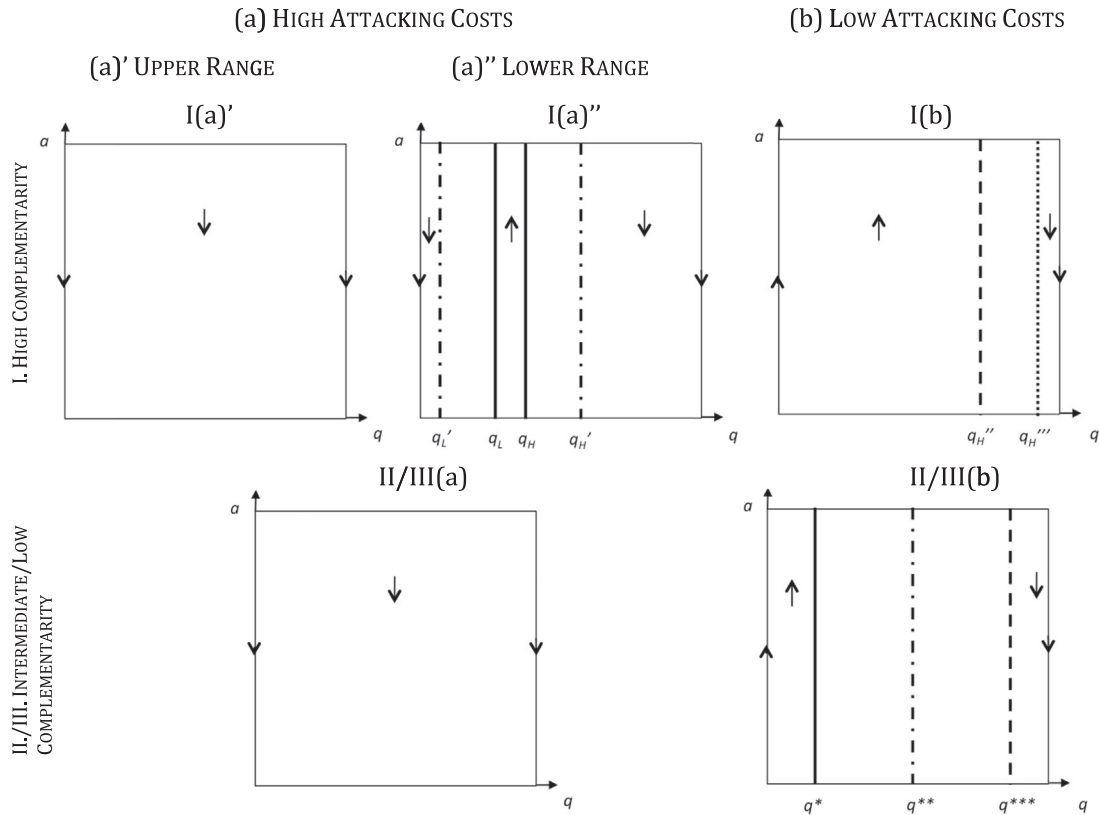


Fig. 5. Position of non-trivial attacker nullclines in the phase space with the fraction of attackers attacking twice on the Y-axis (a) and the fraction of cooperating defenders on the X-axis (q); arrows indicate direction in which attacker population evolves. In the high-complementarity scenario (I), for the upper range of high attacking costs (I(a)'), no non-trivial attacker nullclines exist; for the lower range of high attacking costs (I(a)''), two non-trivial attacker nullclines $q = q_L$ and $q = q_H$ exist, where a increases between these nullclines and decreases otherwise. $q = q_L$ and $q = q_H$ represents attacker nullclines after a decrease in attacking costs. For low attacking costs (I(b)), a single attacker nullcline $q = q^*$ exists, to the left (right) of which a increases (decreases). $q = q^*$ and $q = q^{***}$ represent attacker nullclines after consecutive decreases in attacking costs. In the intermediate- and low-complementarity scenarios (II./III.), the analysis is similar, except that there is no longer an in-between case with two non-trivial attacker nullclines.

low attacking costs, the attacker nullcline $q = q^*$ becomes relevant, where q^* moves from 0 to 1 as we further decrease attacking costs (as illustrated by the consecutive nullclines $q = q^{**}$ and $q = q^{***}$ in part II/III(b) of Fig. 5).

Lemma 2. Endogenous adversity: position in phase space of non-trivial attacker nullclines.

I. High complementarity:

(a)' High attacking costs, upper range: no nullcline within range, attacking once always better.

(a)'' High attacking costs, lower range: there exist two nullclines $q = q_L$ and $q = q_H$ with $0 < q_L < q_H < 1$. One attack is best for $0 \leq q < q_L$; two attacks is best for $q_L < q < q_H$; one attack is again best for $q_H < q \leq 1$. q_L approaches zero as attacking costs approach the lowest level within this cost range, and q_L and q_H both approach the same intermediate level of

q as attacking costs approach the highest level within this cost range.

(b) Low attacking costs: there exists a single nullcline $q = q_H$ with $0 < q_H < 1$. In this case, two attacks is best for $0 \leq q < q_H$; one attack is best for $q_H < q \leq 1$. Within this case, q_H approaches one as attacking costs approach zero, and q_H approaches an intermediate level as attacking costs approach the highest level within this cost range.

II./III. Intermediate/low complementarity:

(a) High attacking costs: no nullcline within range, attacking once always better.

(b) Low attacking costs: there exists a single nullcline $q = q^*$ with $0 < q^* < 1$. In this case, two attacks is best for $0 \leq q < q^*$; one attack is best for $q^* < q \leq 1$. Within this case, q^* approaches one as attacking costs approach zero, and q^* approaches zero as attacking costs approach the highest level within this cost range.

Proof: Appendix A.

Having derived the position of the non-trivial nullclines, we are now ready to list the fixed points of the system (1) and (2). It is obvious that there are three types of fixed points. First, since the edges of the phase space are also nullclines, the four vertices of the phase space, where neither of the populations is mixed, are always fixed points. Second, when the non-trivial defender nullcline ($f_C - f_D = 0$) intersects the trivial attacker nullcline $a = 1$, a fixed point exists with a mixed defender population and a non-mixed attacker population. It is clear from Fig. 3 that such a fixed point always exists for the high-complementarity scenario, and additionally exists for the case of high cooperation costs within the intermediate- and low-complementarity scenarios. Third, when a non-trivial attacker nullcline as characterized in Lemma 2 intersects with a non-trivial defender nullcline as characterized in Lemma 1, an interior fixed point exists with two mixed populations.

The exact conditions under which an interior fixed point exists are given in Appendix B, but the existence of such fixed points in part becomes clear when combining Figs. 3 and 5. By combining part I(a)' of Fig. 5 with parts I(i) or I(ii) of Fig. 3, and part II/III(a) of Fig. 5 with parts II(i), II(ii), III(i), or III(ii) of Fig. 3, it is clear that for high attacking costs, there is no interior fixed point, simply because attackers always obtain higher fitness when attacking once. For high complementarity, an exception is the combination of part I(a)' of Fig. 5 (lower range of high attacking costs) with parts I(i) (low cooperation costs) and I(ii) (high cooperation costs) of Fig. 3. It is clear from Figs. 3 and 5 that an interior fixed point is possible for these cases; Appendix B derives for which attacking cost levels such fixed points exist.

In the intermediate-/low-complementarity scenarios, first, combining low attacking costs in part II/III(b) of Fig. 5, with the case of low cooperation costs in parts II(i) and III(i) of Fig. 3, it is clear that an interior fixed point always exists in these cases. Second, combining low attacking costs in part II/III(b) of Fig. 5, with the case of high cooperation costs (parts II(ii) and III(ii) of Fig. 3), it is clear that an interior fixed point may exist, and Appendix B derives for what attacking cost ranges such fixed points exist. Specifically for the low-complementarity scenario, by the fact that the unique attacker nullcline in part II/III(b) of Fig. 5 shifts to the right as attacking costs are decreased, and by the shape of the defender nullcline in part III(ii) of Fig. 3, it is clear that an interior fixed point exists for the upper range of attacking costs within the case of low attacking costs, whereas no such fixed point exists for the lower attacking costs within this case.

Appendix B also looks at the asymptotic stability of the fixed points. First, among the vertices, the fixed point where all defenders cooperate and all attackers attack twice cannot be asymptotically stable. As the full value of the common resource is now maintained whether or not an attacker attacks twice, and as attacking is costly, the result follows directly. Different from the case with endogenous adversity, there is therefore no stable state where all defenders cooperate and two attacks take place. Each of the other three fixed points is stable for specific ranges of the attacking and the cooperation costs. For high attacking costs (parts I(a)', I(a)'' and II/III(a) of Fig. 5), attackers are better off attacking once both when all defenders cooperate and when they all

defect, so that the fixed points where all attackers attack twice cannot be stable. The stable fixed point here is found by looking at the direction of the trajectory coinciding with the edge $a = 0$ in Fig. 3, and is exactly the same as in the case where the number of attacks is exogenously fixed at one (see Proposition 1); the fixed point where all defenders defect is stable, except in the case with low cooperation costs of the high-complementarity scenario.

For low attacking costs (parts I(b) and II/III(b) of Fig. 5), attackers are better off attacking twice when all defenders defect, and the fixed point where all attackers attack once and all defenders defect cannot be stable. As is clear from Fig. 3, the fixed point where all attackers attack twice and all defenders defect, is now stable in the high-complementarity scenario (parts I(i) and I(ii) of Fig. 3), and in the case with high cooperation costs of the intermediate-complementarity scenario (part II(ii) of Fig. 3), and is not stable otherwise. Part I(i) of Fig. 3 also makes it clear that in the case with low cooperation costs of the high-complementarity scenario, the fixed point where all defenders cooperate and all attackers attack once is additionally stable; in all other cases with low attacking costs, the vertices where all attackers attack once are never stable. Finally, in the intermediate-complementarity scenario, the combination of part II/III(b) of Fig. 5 with part II(i) of Fig. 3 reveals that no vertex is stable in the case with low cooperation costs.

Let us next consider the asymptotic stability of the second type of fixed points, where all attackers attack twice and where the population of defenders contains both cooperators and defectors. As is clear from Fig. 3, in the high-complementarity scenario (parts I(i) and I(ii) of Fig. 3), and in the case with high cooperation costs of the intermediate-complementarity scenario (part II(ii) of Fig. 3), these fixed points are not stable, as the defender game is then a Stag Hunt. For a Stag Hunt, at any such a fixed point, a small increase (decrease) in the fraction of cooperators means that the fraction of cooperators increases (decreases). Furthermore, it is clear from Fig. 3 that in the case with high cooperation costs of the scenario with low-complementarity (part III(ii) of Fig. 3), the defender game when all attackers attack twice is a Snowdrift game. For sufficiently low attacking costs, where no interior fixed point exists, at the fixed point of the second type, attackers are better off attacking twice. Moreover, in the defender game for the number of attacks fixed at two, a small increase (decrease) in the fraction of cooperating defenders means that the fraction of cooperating defenders decreases (increases). It follows that the fixed point of the second type is stable in this case.

Finally, let us consider the asymptotic stability of any fixed point of the third type, where both populations are mixed (interior fixed point). Using linear stability analysis, we show in Appendix B that any such fixed point is not asymptotically stable when $k > 1/2$ (high- or intermediate-complementarity scenario), and is asymptotically stable for $k < 1/2$ (low-complementarity scenario). Intuitively, for $k > 1/2$, the defender game has Stag Hunt features; in such a game, having a large fraction of cooperators makes cooperating more attractive, and having a large fraction of defectors makes defecting more attractive, explaining why trajectories point towards the boundary of the phase space. For $k < 1/2$, the defender game has Snowdrift features; in such a game, having a large fraction of cooperators makes defecting more attractive, and having a large fraction of defectors makes cooperating more attractive, explaining why trajectories point towards the fixed point.

Our results for the model with endogenous adversity, where we only consider asymptotically stable fixed points, are summarized in Proposition 2. The following cases can be distinguished. In the case with low cooperation costs of the high-complementarity scenario (Case I(i)), with high attacking costs, attackers attack once, and all defenders cooperate in the asymptotically stable state. For low attacking costs, as the defender game is a Stag Hunt for a number of attacks fixed at two, the only stable state when attackers attack twice is one where all defenders defect. At the same time, the fixed point where all defenders cooperate and the attackers attack once continues to be stable. As it is

possible that joint defection evolves with low attacking costs, the competing effect continues to apply here.

In the case with high cooperation costs of the high-complementarity scenario (Case I(ii)), with high attacking costs, in the asymptotically stable state attackers attack once, and all defenders defect. For low attacking costs, the defender game is again a Stag Hunt for a number of attacks fixed at two, but the only stable state is one where all defenders defect and attackers attack twice. It follows that, contrary to what is the case in Proposition 1, the common-enemy effect no longer applies here. While joint cooperation can evolve for a number of attacks fixed at two, once joint cooperation is achieved, attackers evolve to set one attack, which again makes joint defection evolve among the defenders, fixing the populations at a stable point. Intuitively, attackers are shielded from the common-enemy effect because once a high attacking level induces defenders to cooperate, attackers evolve to lower the number of attacks, thus undoing the common-enemy effect.

In each of the remaining cases, for high attacking costs the asymptotically stable state is always one where all defenders defect and where all attackers attack once, and the focus is therefore on what happens with low attacking costs. In the case with low cooperation costs of the intermediate-complementarity scenario (Case II(i)), with low attacking costs, no vertex is stable, and the interior fixed point is not stable either. There is therefore no stable fixed point, suggesting a stable heteroclinic cycle, as can be seen by the arrows on the edges of part II(i) of Fig. 3. Intuitively, attackers are shielded from the common-enemy effect for the same reason as in Case I(ii). Yet, because the defender game now is a Harmony game in case of two attacks, the populations cannot settle at a fixed point with two attacks and joint defection by the defenders, because it is not stable. Thus, after defenders have achieved joint cooperation in response to a high attacking level, and after attackers have evolved to reduce the number of attacks, joint defection again evolves among defenders, after which attackers again evolve to set a high number of attacks.

In the case with high cooperation costs of the intermediate-complementarity scenario (Case II(ii)), the game for a number of attacks fixed at two is again a Stag Hunt, and the results are identical to those of Case I(ii). In the case with low cooperation costs of the low-complementarity scenario (Case III(i)), with low attacking costs, the game for a number of attacks fixed at two is a Harmony game, and just as in Case II(i), none of the vertices are asymptotically stable. Yet, because of the low complementarity, the defender game when the attacker population is mixed has Snowdrift features, explaining why the interior fixed point is stable. The common-enemy effect applies here, as a switch from high to low attacking costs means moving from a situation where all defenders defect, to one where a fraction of defenders cooperate. In the case with high cooperation costs of the low-complementarity scenario (Case III(ii)), with low attacking costs, the game for a number of attacks fixed at two is itself a Snowdrift game. As previously derived, when an interior fixed point does not exist, the fixed point of the second type is stable. When a fixed point of the second type does not exist, an interior fixed point exists, and is stable, again because the defender game has Snowdrift features. In both cases, the common-enemy effect again applies.

Proposition 2. *Endogenous adversity: asymptotically stable fixed points, and implications for the relevance of the common-enemy effect/competing effect.*

I. High complementarity:

- (i) Low cooperation costs: \Rightarrow competing effect.
 - (a) High attacking costs: $a=0, q=1$.
 - (b) Low attacking costs: $a=0, q=1$, or $a=1, q=0$.
- (ii) High cooperation costs: \Rightarrow no common-enemy effect
 - (a) High attacking costs: $a=0, q=0$.
 - (b) Low attacking costs: $a=1, q=0$.

II. Intermediate complementarity:

- (i) Low cooperation costs: \Rightarrow partially, common-enemy effect
 - (a) High attacking costs: $a=0, q=0$.
 - (b) Low attacking costs: no stable fixed point.
- (ii) High cooperation costs: \Rightarrow no common-enemy effect
 - (a) High attacking costs: $a=0, q=0$.
 - (b) Low attacking costs: $a=1, q=0$.

III. Low complementarity:

- (i) Low cooperation costs: \Rightarrow common-enemy effect
 - (a) High attacking costs: $a=0, q=0$.
 - (b) Low attacking costs: mixing with $0 < a < 1, 0 < q < 1$.
- (ii) High cooperation costs: \Rightarrow common-enemy effect
 - (a) High attacking costs: $a=0, q=0$.
 - (b) Low attacking costs: mixing with $a=1, 0 < q < 1$ for lower attacking costs within this range; mixing with $0 < a < 1, 0 < q < 1$ for higher attacking costs within this range.

Proof: see Appendix B.

The analysis in Proposition 2 focuses on asymptotically stable fixed points, and ignores the possibility of limit cycles. In Proposition 3 we show that k serves as a bifurcation parameter, with bifurcation value $k=1/2$. At $k=1/2$, the interior fixed point is a center, with all trajectories in the phase space taking the form of neutrally stable closed orbits cycling around this center. It follows that system (1) and (2) undergoes a Hopf bifurcation that is neither subcritical nor supercritical at $k=1/2$. This shows that for k just above $1/2$, all trajectories spiral away from the interior fixed point (which is an unstable spiral), whereas for k just below $1/2$, all trajectories spiral towards the interior fixed point (which is a stable spiral), meaning that in the neighborhood of $k=1/2$, there are no limit cycles. This result does not allow us to draw conclusions about levels of k further away from $k=1/2$. Yet, the fact that no other bifurcations occur for other levels of k , suggests that there are no limit cycles in general.

Proposition 3. Exactly in between intermediate and low complementarity ($k=1/2$), for low attacking costs and high cooperation costs, the interior fixed point is a center, with neutrally stable closed orbits around it. It follows that k functions as a bifurcation parameter, with $k=1/2$ the bifurcation value. The bifurcation is neither subcritical nor supercritical, meaning that for k just above $1/2$, the interior fixed point is an unstable spiral with all trajectories spiraling away from it towards the edges of the phase space; for k just below $1/2$, the interior fixed point is a stable spiral with all trajectories spiraling towards it.

Proof: see Appendix C.

5. Discussion

The common-enemy hypothesis of by-product mutualism appeals to the common-sense idea that adversity encourages cooperation (Kropotkin, 1902). Yet, if it is not the physical environment that plays the role of the common enemy, but instead an organism that is itself subject to evolution, one may expect that the enemy evolves to keep the level of adversity low, thus preventing any common-enemy effect from becoming operative. Our model, where defenders defend a common resource against an attacker, confirms this intuition when increased adversity turns the defenders' game into a Stag Hunt. In this case, increased adversity both makes it possible that joint defense evolves, and joint failure to defend. Yet, if joint defense evolves, it becomes less attractive for attackers to attack; attackers evolve to attack less often, thus shielding themselves against the common-enemy effect. When increased adversity instead turns the defenders' game into a Snowdrift game, adversity makes only a fraction of the defenders cooperate. Attackers now individually obtain highest fitness from setting a high number of attacks, and cannot avoid the common-enemy effect.

Our model may explain why observed instances of cooperation often have Snowdrift features, where part of the organisms do the work, and others lag (see e.g. Packer, 1977; Packer and Ruttan, 1988;

Heinsohn and Packer, 1995). Our explanation of this observation is that in cooperative instances that have Stag Hunt features, the common-enemy effect cannot take effect, and cooperation is simply not observed. At the same time, our model predicts that cooperation with Stag Hunt features is still possible when the role of the common enemy is played by the physical environment, and not by a strategic player.

We end by pointing out some possible directions for future research. First, as many instances of cooperation may involve more than two players, the model should be extended to more than two players. For a switch from attacking costs sufficiently high to make one of the strategies dominant, to very low attacking costs, we do not expect the results to differ, though the effect of small changes in the attacking

costs may become non-monotonous, in line with the non-monotonous effect of exogenous adversity recently found by De Jaegher (2017) for multi-player games. Second, adversity may itself be the product of cooperation, where attackers are only able to productively attack if they cooperate. The interaction between groups of cooperating attackers and defenders may be the subject of future research.

Acknowledgements

I would like to thank one referee for helpful and constructive comments. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Appendix A. Proofs of Lemmata 1 and 2

Proof of Lemma 1.

The average fitness of cooperating f_C is calculated in (6) as follows. A defender who cooperates obtains payoffs in the top rows of payoff matrices (4) or (5). With probability q , the defender is paired to a cooperating defender, in which case fitness V is obtained independently of the number of attacks. With probability $(1 - q)$, the cooperating defender is paired with a defecting defender, in which case with probability $(1 - a)$ he faces one attack and obtains the top right payoff in (4), and with probability a he faces two attacks and obtains the top right payoff in (5).

$$f_C = qV + (1-q)[(1-a)(1 - \frac{1}{2}k)V + a(1 - \frac{3}{4}k)V] - c \quad (6)$$

The average fitness of defecting f_D is calculated in a similar way using the bottom rows of matrices (4) and (5):

$$f_D = q[(1-a)(1 - \frac{1}{2}k)V + a(1 - \frac{3}{4}k)V] + (1-q)[(1-a)(1-k)V + a\frac{1}{2}(1-k)V] \quad (7)$$

Subtracting the right-hand sides of (6) and (7), it follows that

$$f_C - f_D = \frac{1}{2}[(1-a)kV + a(1 - \frac{1}{2}k)V + qa(2k-1)V] - c \quad (8)$$

Using (8), consider the non-trivial defender nullcline, such that $(f_C - f_D) = 0$.

1. We first look at the scenario $k > \frac{1}{2}$ (I. high and II. intermediate complementarity in Definition 1).

1A. When $\frac{1}{2}kV < c < \frac{3}{4}kV$ (which by Definition 1 includes the case of high cooperation costs in the scenario $k > \frac{2}{3}$ (case I(ii) in Definition 1), and all considered cooperation costs in the scenario $\frac{1}{2} < k < \frac{2}{3}$ (cases II(i) and II(ii) in Definition 1)), putting the right-hand side of (8) equal to zero and solving for a , the defender nullcline can be expressed as $a = \frac{2c - kV}{[1 - (3/2)k]V + q[2k - 1]V}$, which is decreasing in q , and intersects the edge $q = 1$ at $a = \frac{2c - kV}{\frac{1}{2}kV}$ (see parts I(ii), II(i) and II(ii) of Fig. 3). Note that $\frac{2c - kV}{\frac{1}{2}kV} > 1$ when $c > \frac{3}{4}kV$; given that the defender nullcline is decreasing, it follows that for $c > \frac{3}{4}kV$, it lies outside of the phase space. For the scenario $\frac{1}{2} < k < \frac{2}{3}$ (intermediate complementarity), in the case with $\frac{1}{2}kV < c < (\frac{1}{2} - \frac{1}{4}k)V$ (low cooperation costs, Definition 1 II(i)), the defender nullcline intersects the edge $q = 0$ at $a = \frac{2c - kV}{[1 - (3/2)k]V}$ (see part II(i) of Fig. 3); for the remaining cases I(ii) and II(ii), the defender nullcline does not intersect the edge $q = 0$. For the scenario $\frac{1}{2} < k < \frac{2}{3}$, in the case with $(\frac{1}{2} - \frac{1}{4}k)V < c < \frac{3}{4}kV$ (high cooperation costs, Definition 1 II(ii)), and for the scenario $k > \frac{2}{3}$ (high complementarity, Definition 1 I(ii)), the defender nullcline instead intersects the edge $a = 1$ at $q = \frac{2c - (1 - \frac{1}{2}k)V}{(2k - 1)V}$ (see part II(ii) of Fig. 3).

1B. For the scenario $k > \frac{2}{3}$, in the case with $(\frac{1}{2} - \frac{1}{4}k)V < c < \frac{1}{2}kV$ (high complementarity, low cooperation costs; Definition 1 I(i)), the defender nullcline can be expressed as $a = \frac{kV - 2c}{[(3/2)k - 1]V - q[2k - 1]V}$, which is increasing in q . This nullcline does not intersect the edge $q = 1$. It intersects the edge $q = 0$ at $a = \frac{kV - 2c}{[(3/2)k - 1]V}$, and intersects the edge $a = 1$ at $q = \frac{2c - (1 - \frac{1}{2}k)V}{(2k - 1)V}$ (see part I(i) of Fig. 3). Note that if instead $c < (\frac{1}{2} - \frac{1}{4}k)V$, it is the case that $\frac{2c - (1 - \frac{1}{2}k)V}{(2k - 1)V} < 0$, and given that the defender nullcline is decreasing, it does not intersect the phase space.

2. We next look at the scenario $k < \frac{1}{2}$. In this case, the defender nullcline $(f_C - f_D) = 0$ can be expressed as $a = \frac{2c - kV}{[1 - (3/2)k]V - q[1 - 2k]V}$, which is increasing in q .

2A. For $\frac{1}{2}kV < c < \frac{3}{4}kV$ (low cooperation costs; Definition 1 III(i)), the defender nullcline intersects the edge $q = 1$ at $a = \frac{2c - kV}{k(V/2)}$. The expression $\frac{2c - kV}{k(V/2)}$ is smaller than one given that $c < \frac{3}{4}kV$. The defender nullcline intersects the edge $q = 0$ at $a = \frac{2c - kV}{[1 - (3/2)k]V}$ (see Fig. 3, part III(i)). The expression $\frac{2c - kV}{[1 - (3/2)k]V}$ is smaller than one if $c < (\frac{1}{2} - \frac{1}{4}k)V$. Given the fact that $\frac{3}{4}kV < (\frac{1}{2} - \frac{1}{4}k)V$ for $k < \frac{1}{2}$, $c < (\frac{1}{2} - \frac{1}{4}k)V$ is always valid for the case.

2B. For $\frac{3}{4}kV < c < (\frac{1}{2} - \frac{1}{4}k)V$ (high cooperation costs; Definition 1 III(ii)), the defender nullcline does not intersect the edge $q = 1$. It intersects the edge $q = 0$ at $a = \frac{2c - kV}{[1 - (3/2)k]V}$. The expression $\frac{2c - kV}{[1 - (3/2)k]V}$ is smaller than one given that $c < (\frac{1}{2} - \frac{1}{4}k)V$, and larger than zero given that $c > \frac{1}{2}kV$. Also, the defender nullcline intersects the edge $a = 1$ at $q = \frac{[1 - \frac{1}{2}k]V - 2c}{[1 - 2k]V}$ (see Fig. 3, part III(ii)). The expression $\frac{[1 - \frac{1}{2}k]V - 2c}{[1 - 2k]V}$ is larger than zero given that $c < (\frac{1}{2} - \frac{1}{4}k)V$, and smaller than one given that $c > \frac{3}{4}kV$. □

Proof of Lemma 2. Given a population of defenders where a ratio of q defenders cooperate, the attacker's expected fitness of attacking once equals:

$$f_{A=1} = q^2(-V) + 2q(1-q)[-(1 - \frac{1}{2}k)V] + (1-q)^2[-(1-k)V] - g \quad (9)$$

Expression (9) is explained as follows, taking into account that the attacker obtains the negative of the value of the common resource in the pair to which she is matched. With probability q^2 , the attacker samples two cooperating defenders from the defender population, and they obtain value V from the common resource. With probability $2q(1 - q)$, the attacker samples a cooperating and a defecting defender from the defender population,

and following payoff matrix (4), these defenders on average obtain fitness $(1 - \frac{1}{2}k)V$ from the resource. Finally, with probability $(1 - q)^2$, the attacker samples two defecting defenders, who by the bottom right expression in (4) on average obtain fitness $(1 - k)V$ from the common resource. Using (5), the attacker's expected fitness of attacking twice can be calculated in a similar way:

$$f_{A=2} = q^2(-V) + 2q(1-q)[-(1 - \frac{3}{4}k)V] + (1-q)^2[-\frac{1}{2}(1-k)V] - 2g \quad (10)$$

Subtracting (9) from (10), we find that:

$$f_{A=2} - f_{A=1} = (1-q)\frac{1}{2}V[q(2k-1) + 1-k] - g \quad (11)$$

The sign of (11), and its roots, determine whether or not the attacker population changes, and when it does, in which direction it changes. Note that $(f_{A=2} - f_{A=1})|_{q=0} = \frac{1}{2}(1-k)V - g$, and $(f_{A=2} - f_{A=1})|_{q=1} = -g < 0$. Furthermore, note that $\partial(f_{A=2} - f_{A=1})/\partial q = -qV(2k-1) + \frac{1}{2}V(3k-2)$, so that $\partial(f_{A=2} - f_{A=1})/\partial q|_{q=0} \geq 0$ when $k \geq \frac{2}{3}$, and so that $\partial(f_{A=2} - f_{A=1})/\partial q|_{q=1} < 0$. Finally, note that $\partial^2(f_{A=2} - f_{A=1})/\partial q^2 = -V(2k-1) \leq 0$ when $k \geq \frac{1}{2}$. It follows that $(f_{A=2} - f_{A=1})$ is a concave function of q for $k > \frac{1}{2}$, and a convex function for $k < \frac{1}{2}$. As the right-hand side of (11) is a quadratic equation, it has up to two real roots q_L and q_H over the range $0 \leq q \leq 1$, which define up to two nullclines $q = q_L$ and $q = q_H$. Given this information, we are now able to distinguish different cases.

Consider first the scenario $k > \frac{2}{3}$. When $g < \frac{1}{2}(1-k)V$, as $(f_{A=2} - f_{A=1})|_{q=0} > 0$ and as $(f_{A=2} - f_{A=1})$ is concave, we obtain that $q_L < 0 < q_H < 1$. It follows that $A = 2$ is better than $A = 1$ for $q < q_H$, while the opposite is obtained for $q > q_H$. When $g > \frac{1}{2}(1-k)V$, as $(f_{A=2} - f_{A=1})|_{q=0} < 0$, as $\partial(f_{A=2} - f_{A=1})/\partial q|_{q=0} > 0$, and as $(f_{A=2} - f_{A=1})$ is concave, the maximum of $(f_{A=2} - f_{A=1})$ is reached at $q = \frac{1}{2}(3k-2)/(2k-1)$, for which $(f_{A=2} - f_{A=1})$ equals $\frac{1}{2}V[\frac{1}{2}k]^2/(2k-1) - g$, where it can be checked that $\frac{1}{2}V[\frac{1}{2}k]^2/(2k-1) > \frac{1}{2}(1-k)V$. Therefore, when $(1-k)(V/2) < g < \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$, we obtain that $0 < q_L < q_H < 1$. It follows that $A = 2$ is now better than $A = 1$ for $q_L < q < q_H$, whereas the opposite is the case for both $q < q_L$ and for $q > q_H$. When instead $g > \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$, $(f_{A=2} - f_{A=1})$ does not have any real roots, and $A = 1$ is better than $A = 2$ for all q .

Next, let us look at the scenario $\frac{1}{2} < k < \frac{2}{3}$. When $g < \frac{1}{2}(1-k)V$, the analysis is analogous to the case $k > \frac{2}{3}$, and we obtain that $q_L < 0 < q_H < 1$, with analogous conclusions about the relative fitness of $A = 2$ and $A = 1$ as a function of q . When instead $g > \frac{1}{2}(1-k)V$, as $(f_{A=2} - f_{A=1})|_{q=0} < 0$, as $\partial(f_{A=2} - f_{A=1})/\partial q|_{q=0} < 0$, and as $(f_{A=2} - f_{A=1})$ is concave, we obtain that $(f_{A=2} - f_{A=1})$ does not have real roots, and $A = 1$ is again better than $A = 2$ for all q .

Finally, let us look at the scenario $k < \frac{1}{2}$. For $g < \frac{1}{2}(1-k)V$, as $(f_{A=2} - f_{A=1})|_{q=0} > 0$, as $(f_{A=2} - f_{A=1})$ is convex and as $(f_{A=2} - f_{A=1})|_{q=1} < 0$ and $\partial(f_{A=2} - f_{A=1})/\partial q|_{q=1} < 0$, we obtain that $0 < q_L < 1 < q_H$. It follows that $A = 2$ is better than $A = 1$ for $q < q_L$, while the opposite is the case for $q > q_L$. For $g > \frac{1}{2}(1-k)V$, as $(f_{A=2} - f_{A=1})|_{q=0} < 0$, as $(f_{A=2} - f_{A=1})$ is convex and as $(f_{A=2} - f_{A=1})|_{q=1} < 0$ and $\partial(f_{A=2} - f_{A=1})/\partial q|_{q=1} < 0$, we obtain that $(f_{A=2} - f_{A=1})$ does not have real roots, and $A = 1$ is again better than $A = 2$ for all q .

To see how the specified roots change as a function of g , and to see which values the roots take for cost levels exactly in between the specified cost ranges, rewrite (11) as the quadratic equation:

$$-\frac{1}{2}V(2k-1)q^2 + \frac{1}{2}V(3k-2)q + \frac{1}{2}V(1-k) - g = 0 \quad (12)$$

to obtain the roots:

$$q_{L,H} = \frac{(3k-2)}{2(2k-1)} \pm \frac{\sqrt{(\frac{1}{2}V)^2 k^2 - 2V(2k-1)g}}{V(2k-1)} \quad (13)$$

where $\frac{(3k-2)}{2(2k-1)}$ is the q for which (12) reaches an extremum; this q is larger than one for $k < \frac{1}{2}$, negative for $\frac{1}{2} < k < \frac{2}{3}$, and lies between 0 and $\frac{1}{2}$ for $k > \frac{2}{3}$. From (13), it follows that q_L decreases in g and q_H increases in g for $k < \frac{1}{2}$, whereas exactly the opposite is true for $k > \frac{1}{2}$. When g approaches zero, $q_{L,H} = \frac{(3k-2)}{2(2k-1)} \pm \frac{k}{2(2k-1)}$, so that for $k > \frac{1}{2}$, we obtain that $q_L = -\frac{(1-k)}{(2k-1)} < 0$, $q_H = 1$; for $k < \frac{1}{2}$, we obtain that $q_L = 1$, $q_H = \frac{(1-k)}{(1-2k)} > 1$. When $g = \frac{1}{2}(1-k)V$, for $k > \frac{2}{3}$, $q_L = 0$ and $q_H = \frac{(3k-2)}{(2k-1)} < 1$; for $\frac{1}{2} < k < \frac{2}{3}$, $q_L = -\frac{(2-3k)}{(2k-1)} < 0$ and $q_H = 0$; for $k < \frac{1}{2}$, $q_L = 0$ and $q_H = \frac{(2-3k)}{(1-2k)} > 1$. For $k > \frac{1}{2}$, when $g = \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$, it is the case that $q_{L,H} = \frac{(3k-2)}{2(2k-1)}$; it follows that for $\frac{1}{2} < k < \frac{2}{3}$, $q_{L,H} = -\frac{(2-3k)}{2(2k-1)} < 0$, whereas for $k > \frac{2}{3}$, $q_{L,H} = \frac{(3k-2)}{2(2k-1)}$, such that $0 < q_{L,H} < \frac{1}{2}$. □

Appendix B. Lemmata for proving Proposition 2

To prove Proposition 2, we show four lemmata from which the proposition directly follows. We first derive the conditions under which the different types of fixed points exist. The existence of the first type of fixed points, namely the vertices, is evident, so that the focus in Lemma A1 lies on fixed points of the second and third types.

Lemma A1. Endogenous adversity: existence of fixed points other than vertices.

I. High complementarity:

(i) Low cooperation costs:

(a) High attacking costs: fixed point $a=1$, $0 < q < 1$; additionally, for a g^* with $\frac{1}{2}V[\frac{1}{2}k]^2/(2k-1) < g^* < kV$, an interior fixed point exists for any g such that $\frac{1}{2}V[\frac{1}{2}k]^2/(2k-1) < g < g^*$.

(b) Low attacking costs: fixed point $a=1$, $0 < q < 1$.

(ii) High cooperation costs:

When $[(5/4)k - \frac{1}{2}]V < c < \frac{3}{4}kV$:

(a) High attacking costs: fixed point $a=1$, $0 < q < 1$;

(b) Low attacking costs: fixed point $a=1$, $0 < q < 1$; additionally, for a g_1^{**} with $0 < g_1^{**} < \frac{1}{2}(1-k)V$, an interior fixed point exists for any g such that $0 < g < g_1^{**}$.

When $\frac{1}{2}kV < c < [(5/4)k - \frac{1}{2}]V$:

(a) High attacking costs: fixed point $a=1$, $0 < q < 1$; additionally, for a g_2^* with $\frac{1}{2}(1-k)V < g_2^* < \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$, an interior fixed point exists for any g such that $\frac{1}{2}(1-k)V < g < g_2^*$.

(b) Low attacking costs: fixed point $a=1$, $0 < q < 1$; interior fixed point.

II. Intermediate complementarity:

- (i) Low cooperation costs:
 - (a) High attacking costs: no fixed points other than vertices;
 - (b) Low attacking costs: interior fixed point.
- (ii) High cooperation costs:
 - (a) High attacking costs: fixed point $a=1$, $0 < q < 1$;
 - (b) Low attacking costs: fixed point $a=1$, $0 < q < 1$; additionally, for a g^{***} with $0 < g^{***} < \frac{1}{2}(1-k)V$, an interior fixed point exists for any g such that $0 < g < g^{***}$.

III. Low complementarity:

- (i) Low cooperation costs:
 - (a) High attacking costs: no fixed points other than vertices;
 - (b) Low attacking costs: interior fixed point.
- (ii) High cooperation costs:
 - (a) High attacking costs: fixed point $a=1$, $0 < q < 1$;
 - (b) Low attacking costs: fixed point $a=1$, $0 < q < 1$; additionally, for a g^{****} with $0 < g^{****} < \frac{1}{2}(1-k)V$, an interior fixed point exists for any g such that $g^{****} < g < \frac{1}{2}(1-k)V$.

Proof:

Existence of a fixed point $a = 1$, $0 < q < 1$.

Such a fixed point is obtained whenever the non-trivial nullcline where the fraction of cooperating defenders does not change, intersects the trivial nullcline $a = 1$. By Lemma 1, it now follows directly that such fixed points exist when $k > \frac{2}{3}$, when $\frac{1}{2} < k < \frac{2}{3}$ for $c > \frac{1}{2}kV$, and when $k < \frac{1}{2}$ also for $c > \frac{1}{2}kV$.

Existence of interior fixed points.

Case I (i).

By Lemma 2.I(a), there is no non-trivial attacker nullcline for $g > \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$ (by Definition 2.I, upper range of high attacking costs), and therefore also no interior fixed point. Consider $g = \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$ (by Definition 2.I, attacking costs exactly in between the upper and the lower range of high attacking costs), so that by Lemma 2, there is a unique attacker nullcline $q_{L,H} = \frac{(3k-2)}{2(2k-1)}$. Then by part 1B of the proof of Lemma 1, the intercept of the defender nullcline with the edge $a = 1$ is $q = \frac{2c - (1 - \frac{1}{2}k)V}{(2k-1)V}$. Given that $\frac{2c - (1 - \frac{1}{2}k)V}{(2k-1)V} < \frac{(3k-2)}{2(2k-1)}$ and that by Lemma 1.I(i) the defender nullcline is increasing, there is therefore no interior fixed point for $g = \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$. As by the proof of Lemma 2, q_L increases in g and q_H decreases, it follows that an interior fixed point can only exist for q_L . Consider next $g = \frac{1}{2}(1-k)V$ (by Definition 2.I, attacking costs exactly in between low and high attacking costs), so that by the proof of Lemma 2, $q_L = 0$. Given that q_L increases in g , it follows that there is no interior fixed point for $g < \frac{1}{2}(1-k)V$ (by Definition 2.I, low attacking costs). As by the proof of Lemma 2 q_L decreases in g , it is clear from Fig. 3.I(i) that a g^* exists such that for all g with $g^* < g < \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$, no interior fixed point exists, and for all g with $\frac{1}{2}(1-k)V < g < g^*$, an interior fixed point exists.

Case I (ii).

By the same arguments as for Case I(i), there is no interior fixed point for $g > \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$. Consider again $g = \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$ (by Definition 2.I, attacking costs exactly in between the upper and the lower range of high attacking costs), so that by Lemma 2, there is a unique attacker nullcline $q_{L,H} = \frac{(3k-2)}{2(2k-1)}$. By part 1A of the proof of Lemma 1, the intercept of the defender nullcline with the edge $a = 1$ is again $q = \frac{2c - (1 - \frac{1}{2}k)V}{(2k-1)V}$. Given that this time $\frac{2c - (1 - \frac{1}{2}k)V}{(2k-1)V} > \frac{(3k-2)}{2(2k-1)}$, and that by Lemma 1.I(ii) the defender nullcline is decreasing, there is therefore no interior fixed point for $g = \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$. As by the proof of Lemma 2, q_L increases in g and q_H decreases, it follows that an interior fixed point can only exist for q_H . Consider next g approaching zero, so that by the proof of Lemma 2, the attacker nullclines are $q_L = -\frac{(1-k)}{(2k-1)} < 0$ and $q_H = 1$. As by the proof of Lemma 2, q_H decreases in g , it follows that for a range of small attacking costs, an interior fixed point exists for q_H . Consider finally $g = \frac{1}{2}(1-k)V$, so that by the proof of Lemma 2 the attacker nullclines are $q_L = 0$, $q_H = \frac{(3k-2)}{2(2k-1)}$. Given that by Lemma 1.I(ii) the defender nullcline is decreasing, it follows that there is an interior fixed point now if $\frac{(3k-2)}{2(2k-1)} > \frac{2c - (1 - \frac{1}{2}k)V}{(2k-1)V}$ iff $c < [(5/4)k - \frac{1}{2}]V$. We note now that given that $\frac{2}{3} < k < 1$, it follows that $\frac{1}{2}kV < [(5/4)k - \frac{1}{2}]V < \frac{3}{4}kV$.

We can now distinguish between two cases. For $[(5/4)k - \frac{1}{2}]V < c < \frac{3}{4}kV$, there is no interior fixed point for g in the lower range of high attacking costs ($\frac{1}{2}(1-k)V < g < \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$). For small attacking costs, a g_1^{**} with $0 < g_1^{**} < \frac{1}{2}(1-k)V$ exists such that for g with $0 < g < g_1^{**}$ an interior fixed point exists, whereas for $g_1^{**} < g < \frac{1}{2}(1-k)V$, it does not.

For $\frac{1}{2}kV < c < [(5/4)k - \frac{1}{2}]V$, there is an interior fixed point for g in the entire range of small attacking costs ($0 < g < \frac{1}{2}(1-k)V$). For the lower range of large attacking costs, a g_2^{**} with $\frac{1}{2}(1-k)V < g_2^{**} < \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$ exists such that for g with $\frac{1}{2}(1-k)V < g < g_2^{**}$ an interior fixed point exists, whereas for $g_2^{**} < g < \frac{1}{2}V[\frac{1}{2}k]^2/(2k-1)$, it does not.

Case II (i).

By Lemma 2.II/III(a), for $g > \frac{1}{2}(1-k)V$ (by Definition 2 II/III, high attacking costs), as there is no non-trivial attacker nullcline, there is no interior fixed point. By the proof of Lemma 2, when $g = \frac{1}{2}(1-k)V$, it is the case that $q_L = -\frac{(2-3k)}{(2k-1)} < 0$ and $q_H = 0$. When g approaches zero, it is the case that $q_L = -\frac{(1-k)}{(2k-1)} < 0$, $q_H = 1$. It follows that an interior fixed point is only possible for $q = q_H$. As by the proof of Lemma 2, q_H decreases in g , and as by Lemma 1.II(i), the defender nullcline cuts $q = 0$ and $q = 1$, it follows that an interior fixed point exists of the entire range of low attacking costs.

Case II (ii).

The analysis is similar to Case II(i), with the difference that by Lemma 1.II(ii), the defender nullcline cuts $a = 1$ and $q = 1$. As by the proof of Lemma 2, q_H decreases in g , it follows that a g^{***} with $0 < g^{***} < \frac{1}{2}(1-k)V$ exists such that for g with $0 < g < g^{***}$ the nullclines intersect, whereas for g with $g^{***} < g < \frac{1}{2}(1-k)V$, they do not.

Case III (i).

By Lemma 2.II/III(a), for $g > \frac{1}{2}(1-k)V$ (by Definition 2 II/III, high attacking costs), as there is no non-trivial attacker nullcline, there is again no interior fixed point. When $g = \frac{1}{2}(1-k)V$, by (13), it is the case that $q_L = 0$ and $q_H = \frac{(2-3k)}{(1-2k)} > 1$. When g approaches zero, by (13), it is the case that

$q_L=1$, $q_H = \frac{(1-k)}{(1-2k)} > 1$. It follows that an interior fixed point is only possible for $q = q_L$. So, only q_L is relevant, and ranges from 0 to 1 in the range of low attacking costs. As by the proof of Lemma 2, q_L decreases in g , and as by Lemma 1.III(i), the defender nullcline cuts $q = 0$ and $q = 1$, it follows that an interior fixed point exists of the entire range of low attacking costs.

Case III (ii).

The analysis is similar to Case III(i), with the difference that by Lemma 1.III(ii), the defender nullcline cuts $a = 1$ and $q = 0$. As by the proof of Lemma 2, q_L decreases in g , it follows that a g^{****} with $0 < g^{****} < \frac{1}{2}(1-k)V$ exists such that for g with $0 < g < g^{****}$ the nullclines intersect, whereas for g with $g^{****} < g < \frac{1}{2}(1-k)V$, they do not. \square

We next investigate the asymptotic stability of the three different types of fixed points in separate lemmata, and start with the vertices in Lemma A2.

Lemma A2. The following three vertices are asymptotically stable, under the listed conditions:

- $q=0$, and $a=0$: if $g > \frac{1}{2}(1-k)V$, $c > \frac{1}{2}kV$
- $q=1$, and $a=0$: $c < \frac{1}{2}kV$
- $q=0$, and $a=1$: if $g < \frac{1}{2}(1-k)V$, $c > (\frac{1}{2} - \frac{1}{4}k)V$

Proof:

When $q = 1$, attackers always obtain $-V$. As attacking is costly, attackers now obtain higher fitness when setting $A = 1$, showing that the fixed point $q = 1$, $a = 1$ is unstable.

When $a = 0$, by (8), the individual defender obtains higher fitness from defecting when all other defenders defect if $c > \frac{1}{2}kV$. By (11), there is now a fixed point $a = 0$, $q = 0$ if additionally $g > \frac{1}{2}(1-k)V$. Also by (11), there is now a fixed point $a = 0$, $q = 1$ for any g .

When $a = 1$, $q = 0$, by (8), the individual defender obtains higher fitness from defecting when all other defenders defect if $c > (\frac{1}{2} - \frac{1}{4}k)V$. By (11), attackers now prefer attacking twice if $g < \frac{1}{2}(1-k)V$. \square

Next, we look at the asymptotic stability of the second type of fixed points, where the defender population is mixed with $0 < q < 1$, but where the attacker population is not mixed, and in particular $a = 1$.

Lemma A3.

- (i) For $k > \frac{1}{2}$, any fixed point with $0 < q < 1$ and $a=1$ is not asymptotically stable;
- (ii) For $k < \frac{1}{2}$, in the case with high cooperation costs of the low-complementarity scenario, with attacking costs g such that $0 < g < g^{****}$ (with g^{****} defined in Lemma A1, part III(ii)(b)), the fixed point with $0 < q < 1$ and $a=1$ is asymptotically stable.

Proof:

By (8), $f_C - f_D$ is increasing in q for $k > \frac{1}{2}$. It follows that q decreases (increases) for fractions of cooperating defenders lower (higher) than in any fixed point where $(f_C - f_D) = 0$, so that such a fixed point is necessarily unstable. Instead, $f_C - f_D$ is increasing in q when $k < \frac{1}{2}$, and q increases (decreases) for fractions of cooperating defenders lower (higher) than in such a fixed point. By Lemma A1, under the specified conditions, attackers are additionally better off attacking twice rather than once at the specified fixed point. \square

We finally look at the asymptotic stability of the third type of fixed points, namely interior fixed points.

Lemma A4. If an interior fixed point exists, it is not asymptotically stable for $k > \frac{1}{2}$, and asymptotically stable for $k < \frac{1}{2}$.

Proof:

We perform linear stability analysis around any fixed point of the mentioned form. Using (1), (2), (8) and (11), the Jacobian matrix at this fixed point takes the form:

$$\begin{bmatrix} q(1-q)[\frac{1}{2}a(2k-1)V] & -q(1-q)\frac{1}{4}VX \\ a(1-a)\frac{1}{2}VX & 0 \end{bmatrix} \quad (14)$$

where $X = \{-q2(2k-1) + (3k-2)\}$. It follows that the trace is positive for intermediate or high complementarity ($k > \frac{1}{2}$), and negative for small complementarity ($k < \frac{1}{2}$). Furthermore, the determinant equals $q(1-q)\frac{1}{4}Va(1-a)\frac{1}{2}VX^2$, which is always positive. It follows that any interior fixed point is unstable for $k > \frac{1}{2}$, and is stable for $k < \frac{1}{2}$. \square

Appendix C. Proof of Proposition 3

Proof of Proposition 3. We first separately analyze the case $k=\frac{1}{2}$, which was so far ignored. In this case, as the number of attacks is exogenously increased from one to two (cf. Proposition 1), the game changes from a Prisoner's Dilemma into a Harmony game, where we note it is the case that $\frac{1}{2}kV < c < \frac{3}{4}kV = (\frac{1}{2} - \frac{1}{4}k)V$, meaning that the case of large cooperation costs vanishes. We next look at the position of the non-trivial nullcline $\dot{q}=0$ (cf. Lemma 1). The nullcline reduces to $a = \frac{2c-kV}{[1-(3/2)k]V}$, which lies between 0 and 1 given that $\frac{1}{2}kV < c < (\frac{1}{2} - \frac{1}{4}k)V = \frac{3}{4}kV$.

By (14), the trace is zero for complementarity exactly between intermediate and low ($k=\frac{1}{2}$). Also, the discriminant equals $q^2(1-q)^2(\frac{1}{2}a[2k-1]V)^2 - 4q(1-q)(V/4)a(1-a)\frac{1}{2}VX^2$, which is negative for $k=\frac{1}{2}$, and by continuity negative for k around $\frac{1}{2}$ as well. It follows that any interior fixed point is a center for $k=\frac{1}{2}$, an unstable spiral for k just above $\frac{1}{2}$, and a stable spiral just below k , from which the bifurcation analysis follows.

References

- Archetti, M., Scheuring, I., 2012. Review: game theory of public goods in one-shot dilemmas without assortment. *J. Theor. Biol.* 299, 9–20.
- Arenas, A., Camacho, J., Cuesta, J.A., Requejo, R.J., 2011. The joker effect: cooperation driven by destructive agents. *J. Theor. Biol.* 279, 113–119.
- Brown, J.L., 1983. Cooperation: a biologist's dilemma. In: Rosenblatt, J.S. (Ed.), *Advanced in the Study of Behavior*. Academic Press, New York, 1–37.
- Bruintjes, R., Lynton-Jenkins, J., Jones, J.W., Radford, A.N., 2016. Out-group threat promotes within-group affiliation in a cooperative fish. *Am. Nat.* 187, 274–282.
- Callaway, R.G., Booker, R.W., Choler, P., Kikvidze, Z., Lortie, C.J., Michalet, R., Paolini, L., Pugnaire, F.I., Newingham, B., Aschehoug, E.T., Armas, C., Kikodze, D., Cook, B.J., 2002. Positive interactions among alpine plants increase with stress. *Nature* 417, 844–848.
- De Jaegher, K., 2017. Harsh environments and the evolution of multi-player cooperation. *Theor. Pop. Biol.* 113, 1–12.
- De Jaegher, K., Hoyer, B., 2016a. By-product mutualism and the ambiguous effects of harsher environments – a game-theoretic model. *J. Theor. Biol.* 393, 82–97.
- De Jaegher, K., Hoyer, B., 2016b. Collective action and the common enemy effect. *Def. Peace Econ.* 27, 644–664.
- Doebeli, M., Hauert, C., 2005. Models of cooperation based on the Prisoner's Dilemma and the Snowdrift game. *Ecol. Lett.* 8, 748–766.

- Dugatkin, L.A., 1997. *Cooperation Among Animals: An Evolutionary Perspective*. Oxford University Press, New York.
- Dugatkin, L.A., 2002a. Cooperation in animals: an evolutionary overview. *Biol. Philos.* 17, 459–476.
- Dugatkin, L.A., 2002b. Animal cooperation among unrelated individuals. *Naturwiss* 89, 533–541.
- Gese, E.M., 2001. Territorial defense by coyotes (*Canis latrans*) in Yellowstone National Park, Wyoming: who, how, where, when, and why. *Can. J. Zool.* 79, 980–987.
- Grinnell, J., Packer, C., Pusey, A.E., 1995. Cooperation in male lions: kinship, reciprocity or mutualism? *Anim. Behav.* 49, 95–105.
- Han, T.A., Pereira, L.M., Santos, F.C., 2011. Intention recognition promotes the emergence of cooperation. *Adapt. Behav.* 19, 264–279.
- Heinsohn, R., Packer, C., 1995. Complex cooperative strategies in group-territorial African lions. *Science* 269, 1260–1262.
- Hirshleifer, J., 1983. From weakest-link to best-shot: the voluntary provision of public goods. *Public Choice* 41, 371–386.
- Jolivet, P., Vasconcellos-Neto, J., Weinstein, P., 1990. Cycloalexy: a new concept in the larval defense of insects. *Insect. Mundi* 4, 133–141.
- Krams, I., Berzins, A., Krama, T., Wheatcroft, D., Igaune, K., Rantala, M.J., 2010. The increased risk of predation enhances cooperation. *Proc. R. Soc. B-Biol. Sci.* 277, 513–518.
- Kropotkin, P., 1902. *Mutual Aid*. Heinemann, London.
- Maynard Smith, J., Price, G.R., 1973. The logic of animal conflict. *Nature* 246, 15–18.
- Mesterton-Gibbons, M., Dugatkin, L.A., 1992. Cooperation among unrelated individuals: evolutionary factors. *Q. Rev. Biol.* 67, 267–281.
- Mesterton-Gibbons, M., Dugatkin, L.A., 1997. Cooperation and the prisoner's dilemma: towards testable models of mutualism versus reciprocity. *Anim. Behav.* 54, 551–557.
- Nowak, M.A., 2006. Five rules for the evolution of cooperation. *Science* 314, 1560–1563.
- Packer, C., 1977. Reciprocal altruism in *Papio anubis*. *Nature* 265, 441–443.
- Packer, C., Rutten, L., 1988. The evolution of cooperative hunting. *Am. Nat.* 132, 159–198.
- Polizzi di Sorrentino, E., Schino, G., Massaro, L., Visalberghi, E., Aureli, F., 2012. Between-group hostility affects within-group interactions in tufted capuchin monkeys. *Anim. Behav.* 83, 445–451.
- Port, M., Kappeler, P.M., Johnstone, R.A., 2011. Communal defense of territories and the evolution of sociality. *Am. Nat.* 178, 787–800.
- Ray, D., Baland, J.M., Dagnelie, O., 2007. Inequality and inefficiency in joint projects. *Econ. J.* 117, 922–935.
- Roberts, G., 2005. Cooperation through interdependence. *Anim. Behav.* 70, 901–908.
- Rubenstein, D.I., Nuñez, C.N., 2009. Sociality and reproductive skew in horses and zebras. In: Hager, R., Jones, C.B. (Eds.), *Reproductive Skew in Vertebrates: Proximate and Ultimate Courses*. Cambridge University Press, Cambridge, 196–226.
- Russell, B.W., 1959. *Common Sense and Nuclear Warfare*. George Allen and Unwin, London.
- Sachs, J.L., Mueller, U.G., Wilcox, T.P., Bull, J.J., 2004. The evolution of cooperation. *Q. Rev. Biol.* 79, 135–160.
- Sandoval, L., Wilson, D.R., 2012. Local predation pressure predicts the strength of mobbing responses in tropical birds. *Curr. Zool.* 58, 781–790.
- Scheel, D., Packer, C., 1991. Group hunting behavior of lions: a search for cooperation. *Anim. Behav.* 41, 697–709.
- Skyrms, B., 2004. *The Stag Hunt and Evolution of Social Structure*. Cambridge University Press, Cambridge Ma.
- Spieler, M., 2003. Risk of predation affects aggregation size: a study with tadpoles of *Phrynomantis microps* (Anura: microhylidae). *Anim. Behav.* 65, 179–184.
- Stander, P.E., 1992. Cooperative hunting in lions: the role of the individual. *Behav. Ecol. Socio.* 29, 445–454.
- Strassman, J.E., Zhu, Y., Queller, D.C., 2000. Altruism and social cheating in social amoeba. *Dictyostelium Disco. Nat.* 408, 965–967.
- Sugden, R., 1986. *The Economics of Rights, Cooperation and Welfare*. Blackwell, Oxford.
- Taylor, P.D., Jonker, L.B., 1978. Evolutionarily stable strategies and game dynamics. *Math. Biosci.* 40, 145–156.
- Tucker, A., 1950. A two-person dilemma. In: Rasmussen, E. (Ed.), *Readings in Games and Information*. Blackwell, Oxford, 7–8.
- West Eberhard, M.J., 1975. The evolution of social behavior by kin selection. *Q. Rev. Biol.* 50, 1–35.