



# The evolution of cooperation in finite populations with synergistic payoffs

Rafael Ventura<sup>1</sup>

Received: 20 October 2018 / Accepted: 27 July 2019 / Published online: 2 August 2019  
© Springer Nature B.V. 2019

## Abstract

In a series of papers, Forber and Smead (J Philos 111(3):151–166, 2014, Biol Philos 30(3):405–421, 2015) and Smead and Forber (Evolution 67(3):698–707, 2013) make a valuable contribution to the study of cooperation in finite populations by analyzing an understudied model: the prisoner’s delight. It always pays to cooperate in the one-shot prisoner’s delight, so this model presents a best-case scenario for the evolution of cooperation. Yet, what Forber and Smead find is highly counter-intuitive. In finite populations playing the prisoner’s delight, increasing the benefit of cooperation causes selection to favor defection. Here, I extend their model by considering the effects of non-linear payoffs. In particular, I show that interesting subtleties arise when payoffs are synergistic. Indeed, analysis reveals that increasing the benefit of cooperation does not always favor the spread of defection if payoffs are synergistic. I conclude by drawing some general considerations about robustness analysis in evolutionary models.

**Keywords** Cooperation · Prisoner’s delight · Finite populations · Synergy

## Introduction

Game-theoretic models in biology often assume that evolution takes place in an infinite population. This may be a good approximation of large, panmictic populations. But some biological systems consist of small, finite populations. In finite populations, game-theoretic models have long been known to yield different results than in infinite populations—see Hamilton (1971) for an early discussion and Taylor et al. (2004) for a more recent and comprehensive treatment of finite-population models. Population size has also been shown to affect the evolution of cooperation in well-known settings, such as the prisoner’s dilemma (Nowak et al. 2004). In a series of papers, Forber and Smead (2014, 2015) and Smead and Forber (2013) make a

---

✉ Rafael Ventura  
rhtventura@gmail.com

<sup>1</sup> Department of Philosophy, Bilkent University, Ankara, Turkey

valuable contribution to the study of cooperation in finite populations by analyzing an understudied model: the prisoner's delight. It always pays to cooperate in the one-shot prisoner's delight, so this model presents a best-case scenario for the evolution of cooperative behavior. Yet, what Forber and Smead find is highly counterintuitive: in finite populations playing the prisoner's delight, increasing the benefit of cooperation favors the spread of defection.

Here, I extend their model by considering the effects of non-linear payoffs. In particular, I show that interesting subtleties arise when payoffs are synergistic—that is, when the gain in payoff that agents get from an additional cooperator entering the population depends on the number of cooperators already present in the population. Indeed, analysis of the model reveals that increasing the benefit of cooperation does not always favor the spread of defection if payoffs are synergistic.

To argue for this, I begin by providing a brief description of Forber and Smead's model in “[The prisoner's delight](#)” section. After assessing their model in “[Evaluation of the prisoner's delight](#)” section, I present in “[Prisoner's delight with synergistic payoffs](#)” section an extension of the prisoner's delight in which payoffs are synergistic. I then consider how synergistic payoffs affect the evolutionary fate of cooperation in a finite-population model. Finally, I conclude in “[Conclusion](#)” section by drawing some general considerations about the nature of modeling and robustness analysis in evolutionary biology.

## The prisoner's delight

Forber and Smead consider slightly different versions of the prisoner's delight in different papers. To help understand the counter-intuitiveness of their main result, I focus here mainly on the version of the game that Forber and Smead (2014) discuss. Players are picked at random from a finite population to engage in fitness-affecting interactions. There are two types of player: a player that cooperates (*C*), and a player that defects (*D*). If player *i* is a cooperator, player *j* receives a benefit of *b*; if player *i* is a defector, an amount *a* is deducted from *i*'s payoff and *j* receives a payoff of 1 (with  $b > a > 0$  and  $b > 1$ ). As usual in evolutionary game theory, each type of player is assumed to play a pure strategy that is heritable. Mixed strategies are taken to represent polymorphic states of the population. For simplicity, assume that there is no mutation.

From Table 1, it is easy to see that it always pays to cooperate: defecting always decreases one's payoff. For this reason, an indefinitely large population of cooperators is also an evolutionary stable state. A population of cooperators is evolutionarily stable against invasion by defectors provided that the following conditions are met: (1) cooperators obtain a payoff that is at least as high as that of defectors in interactions with other cooperators, and (2) cooperators obtain a higher payoff than defectors in interactions with defectors if cooperators and defectors obtain the same payoff in interactions with cooperators (Maynard-Smith and Price 1973). In the prisoner's delight, cooperators always obtain a higher payoff than defectors so these two conditions are met. A large population of cooperators therefore cannot be invaded by a small number of defectors.

**Table 1** Payoff matrix for the prisoner’s delight

	C	D
C	$b, b$	$1, b - a$
D	$b - a, 1$	$1 - a, 1 - a$

Rows and columns represent types of players. In each cell, the first entry represents the payoff to the column player; the second entry represents the payoff to the row player. Assume that  $b > 1$  and  $b > a > 0$

This is not necessarily true in finite populations. What is more, Forber and Smead show that increasing the value of  $b$  makes it less likely that cooperation evolves. To see how they derive this seemingly paradoxical result, it is helpful first to define fitness functions for the players. Let  $F_C(\vec{x})$  denote the fitness of cooperators and  $F_D(\vec{x})$  denote the fitness of defectors, where  $x_C$  is the number of cooperators,  $x_D$  is the number of defectors in the population, and  $\vec{x} = \langle x_C, x_D \rangle$  is the composition of the population. The fitness function for each player type is then equal to the mean expected payoff from pairwise interactions. This is equivalent to assuming that each agent plays the entire population at once and that payoffs are a linear function of the number of cooperators and defectors in the population. The fitness functions for cooperators and defectors are given by:

$$F_C(\vec{x}) = \frac{b \cdot (x_C - 1) + x_D}{N - 1} \tag{1a}$$

$$F_D(\vec{x}) = \frac{b \cdot x_C + x_D - 1}{N - 1} - a \tag{1b}$$

These fitness functions assume that the population is finite and composed of  $N$  individuals. Since a cooperator cannot play against herself, there are  $x_C - 1$  cooperators in the remaining population for her to interact with. As a result, to find the mean expected payoff that a cooperator gets from interacting with other cooperators we multiply the benefit of cooperation ( $b$ ) by the number of cooperators in the remaining population ( $x_C - 1$ ). Likewise, each defector can only interact with the remaining  $x_D - 1$  defectors and the fitness function for defectors takes that into account. This is a peculiarity of finite population models: in finite populations, the composition of the remaining population is affected by the removal of the focal individual.

Equipped with these fitness functions, we can now examine the conditions under which selection favors cooperation. Without any mutation, natural selection favors cooperators if their fitness is higher than that of defectors. This condition is given by the following statement:

$$F_C(\vec{x}) > F_D(\vec{x}) \tag{2a}$$

Conversely, defection is favored if the fitness of defectors is higher than that of cooperators. This condition is given by:

$$F_C(\vec{x}) < F_D(\vec{x}) \tag{2b}$$

When  $F_C(\vec{x}) = F_D(\vec{x})$ , both types are equally fit and there is no natural selection at work. Substituting equalities 1a and 1b into expressions 2a and 2b and simplifying, natural selection favors cooperation if the following condition holds:

$$N - 1 > \frac{b - 1}{a} \quad (3a)$$

Similarly, defection is favored if the following condition holds:

$$N - 1 < \frac{b - 1}{a} \quad (3b)$$

These conditions imply that for a large enough value of  $b$ , the fitness of defectors is higher than the fitness of cooperators. Thus, increasing the benefit of cooperative behavior beyond a certain value makes cooperation selectively disadvantageous in a finite population. This is the counterintuitive result that Forber and Smead (2014) call the “spite effect”: in the finite-population prisoner’s delight, defection is a spiteful behavior because it is costly for the defector to perform but under the right conditions it can nonetheless evolve.

This is not the only reason why finite population models pose a problem for the evolution of cooperation. Another reason is what Forber and Smead (2014) call “the nearly neutral effect.” To understand how this effect arises, we must first define the concept of selection coefficient. Selection coefficients measure the strength of natural selection vis-à-vis random drift, which is purely stochastic. The selection coefficient ( $s$ ) is determined by normalizing the fitness functions of competing types. The selection coefficient against defectors is given by:

$$\frac{F_C(\vec{x})}{F_C(\vec{x})} = 1 \quad (4a)$$

$$\frac{F_D(\vec{x})}{F_C(\vec{x})} = 1 - s \quad (4b)$$

The choice of  $F_C(\vec{x})$  as the normalizing factor is arbitrary. Solving for  $s$ :

$$s = \frac{a(N - 1) - b + 1}{b(x_C - 1) + x_D} \quad (5)$$

Notice that increasing the value of  $b$  decreases the value of  $s$ . This does not necessarily favor the spread of defection, but it will decrease the selection pressure against it. Decreasing the strength of selection against defection causes the evolution of cooperation to become more neutral. That is, evolution becomes less subject to natural selection and more subject to random drift.

In sum, Forber and Smead’s analysis correctly leads to the counterintuitive result that increasing the benefit of cooperative behavior can favor the evolution of cooperation in finite populations playing the prisoner’s delight. This is due to the spite

and the nearly neutral effects. Both effects are most pronounced not only when the benefit of cooperation is high, but also when the population is small and the cost of defection is low relative to the benefit of cooperation. In the next section, I offer reasons to extend their model by modifying some of its assumptions.

## Evaluation of the prisoner's delight

As it stands, the model presented by Forber and Smead supports the startling conclusion that increasing the benefit of cooperation makes it less likely that cooperation evolves in finite populations playing the prisoner's delight. It is also instructive, however, to explore what follows when we extend their model and reconsider some of their assumptions. For one, their model assumes a population that evolves according to a Moran process. This means that their model does not specify whether members of the population engage in pairwise interactions, or in multi-way interactions with payoffs that are linear functions of the population composition.

Although the spite and the nearly neutral effects that Forber and Smead identify do not rely on this assumption, in some cases it may be plausible to assume quite explicitly that members of the population engage in multi-way interactions. In both human and nonhuman populations, some cooperative behaviors in nature require the recruitment of several parties. Indeed, the biological world is rife with cooperative behavior of this kind. To name but a few examples, groups with multiple members of female lions often cooperate when hunting—for details, see Packer et al. (1990). Amongst social microbes such as slime models, a multitude of individual cells cooperate during colony formation (Bonner 2009). And in human societies, production of public goods often takes the form of a social dilemma (Hardin 1968). To understand the evolution of cooperation in such cases, it may therefore be helpful to assume explicitly that interactions take place among several members of the population at once.

A more consequential assumption is the payoff structure of the underlying game. Forber and Smead assume that the benefit of cooperation takes on a constant value of  $b$  and that the cost of defection is a constant  $a$ . But it is possible to relax this assumption and suppose that the benefit of cooperation and the cost of defection are nonlinear functions of the number of cooperators and defectors in the population. In fact, cooperative efforts are often synergistic. When few players cooperate, recruiting more cooperators usually has a large effect on the benefit generated. When several players cooperate, adding more cooperators tends to have only a small effect on that benefit. Similarly, how harmful an additional defector is often depends on how many defectors are already present. Synergistic payoffs have indeed been found in a wide range of cooperative group behaviors—for example in cooperative hunting among spiders (Yip et al. 2008), quorum sensing in social bacteria (Miller and Bassler 2001), and perhaps even in the case of human societies working to protect the climate from anthropogenic change (Milinski et al. 2006).

Thus, Forber and Smead's model can be extended so as to incorporate synergistic payoffs and explicitly assume multi-player interactions. In what follows, I present an extension of their model with precisely these features. In keeping with previous

work on the effect of synergy in other game-theoretical models (Allen et al. 2013; Cavalli-Sforza and Feldman 1978; van Veelen 2009), analysis reveals that synergistic payoffs have subtle effects on the evolutionary fate of cooperation in the prisoner's delight.

### Prisoner's delight with synergistic payoffs

Consider the following game. As before, there are  $N$  players in a population of constant size. And there are again two types of player: a type that cooperates ( $C$ ), and a type that defects ( $D$ ). But this time we explicitly assume that all players interact at once. After each round, the population dissolves and individuals reproduce with fitness proportional to their individual payoff. The offspring generation then replaces the parent generation, and the cycle restarts. As before, assume again that defecting is costly. But, most importantly, assume now that the cost of defection and the benefit of cooperation are synergistic functions of the number of players playing  $C$ . This model resembles the scenario underlying the "interdependence hypothesis" for the evolution of human cooperation (Tomasello et al. 2012): small, finite and populations engaging in synergistic interactions in which cooperation is beneficial and defection is costly. For simplicity, assume again that there is no mutation.

Now let the functions  $\alpha(x_C)$  and  $\beta(x_C)$  re-scale the cost of defection and the benefit of cooperation, where  $x_C$  is as before the total number of players playing  $C$  in the entire population ( $0 \leq x_C \leq N$ ). Assume further that the functions  $\alpha(x_C)$  and  $\beta(x_C)$  have the following form:

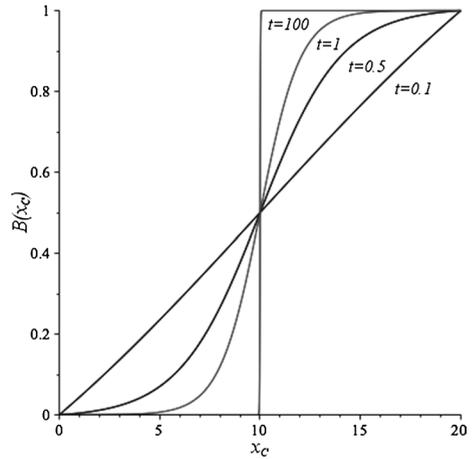
$$\alpha(x_C) = \beta(x_C) = \frac{1}{1 + e^{-t(x_C - k)}} \quad (6)$$

where the parameter  $t$  gives the steepness of the function at the inflection point.<sup>1</sup> The parameter  $t$  therefore determines how smoothly the payoff function changes in response to increases in the number of cooperators around the inflection point: very low values of  $t$  are associated with very smooth changes, with the function behaving much like a linear function as  $t$  approaches zero; moderate values of  $t$  are associated with a sharp change around the inflection point; and very high values of  $t$  are associated with an abrupt switch-like change. Note also that  $k$  is the position of the inflection point along the  $x_C$ -axis. This means that  $k$  is a parameter that controls how many cooperators are required for the payoff function to go from increasing to diminishing returns in the number of cooperators.

For convenience, we restrict  $\alpha(x_C)$  to the range  $[0, a]$  and  $\beta(x_C)$  to the range  $[0, b]$ . This can be done by normalizing and re-scaling  $\alpha(x_C)$  and  $\beta(x_C)$  in the following way:

<sup>1</sup> Strictly speaking, the range of  $\alpha(x_C)$  and  $\beta(x_C)$  is here the set of natural numbers less than  $N$ . So the notions of steepness, inflection point, and other terms that apply to continuous functions are not defined for these functions. Here, I use these terms as a convenient short-hand for what would be the case in the limit of  $\alpha(x_C)$  and  $\beta(x_C)$  as  $N$  goes to infinity, but my results do not depend on this. Similar functions have been studied in non-cooperative settings by Archetti and Scheuring (2011).

**Fig. 1** The benefit of cooperation as a function of the number of cooperators in the population. Shown here is the benefit of cooperation  $B(x_C) = b \frac{\beta(x_C) - \beta(0)}{\beta(N) - \beta(0)}$  for different values of the parameter  $t$ , with  $\beta(x_C) = \frac{1}{1 + e^{-t(x_C - k)}}$ ,  $b = 1$ ,  $k = N/2$ , and  $N = 20$ . Similar results obtain for  $A(x_C)$ , the cost of defection



$$A(x_C) = a \frac{\alpha(x_C) - \alpha(0)}{\alpha(N) - \alpha(0)} \tag{7a}$$

$$B(x_C) = b \frac{\beta(x_C) - \beta(0)}{\beta(N) - \beta(0)} \tag{7b}$$

where  $N$  is the maximal number of players that can choose to cooperate. In a finite population model,  $N$  is also the population size (see Fig. 1).

We can now substitute  $A(x_C)$  and  $B(x_C)$  for the constants  $a$  and  $b$  in the fitness functions given by Eqs. 1a and 1b:

$$F_C(\vec{x}) = \frac{B(x_C) \cdot (x_C - 1) + x_D}{N - 1} \tag{8a}$$

$$F_D(\vec{x}) = \frac{B(x_C) \cdot x_C + x_D - 1}{N - 1} - A(x_C) \tag{8b}$$

These fitness functions allow us to re-state conditions 3a and 3b, which determine when selection will favor cooperation and defection. Without mutation, natural selection will favor cooperators if:

$$N - 1 > \frac{B(x_C) - 1}{A(x_C)} \tag{9a}$$

Conversely, defectors will be favored if:

$$N - 1 < \frac{B(x_C) - 1}{A(x_C)} \tag{9b}$$

The expressions 9a and 9b allow us now to determine the conditions that favor the spread of cooperation and defection. In particular, two extreme cases should be considered in detail: the case in which the parameter  $t$  goes to infinity, and the case in which the parameter  $t$  goes to zero. If  $t \rightarrow \infty$ , then  $A(x_C)$  and  $B(x_C)$  take the form of step functions:

$$A(x_C) = \begin{cases} a & \text{if } x_C \geq k \\ 0 & \text{if } x_C < k \end{cases} \quad (10a)$$

$$B(x_C) = \begin{cases} b & \text{if } x_C \geq k \\ 0 & \text{if } x_C < k \end{cases} \quad (10b)$$

In this limiting case,  $A(x_C)$  and  $B(x_C)$  are both equal to zero when  $x_C < k$ . That is, the benefit of cooperation is null if the number of cooperators in the population falls below a certain threshold ( $k$ ). So when  $x_C < k$ , the right-hand side of 9a and 9b goes to  $-\infty$ , and condition 9b is never satisfied no matter what value the benefit of cooperation ( $b$ ) takes. On the assumption that  $k = N/2$ , this means that selection always favors cooperation whenever less than half of the population consists of cooperators. For lower values of  $k$ , fewer cooperators are required for the selective regime to switch and selection to favor defection. But for higher values of  $k$ , it is harder for selection to favor defection as more cooperators are required to ensure that  $A(x_C)$  and  $B(x_C)$  are not equal to zero and that condition 9b is satisfied. At any rate, the spite effect does not arise when the total number of cooperators falls below the threshold  $k$ .

If  $x_C \geq k$ , on the other hand, then it is possible for condition 9b to be realized. That is, if cooperators already make up a substantial portion of the population, then selection can favor defection depending on how large the cost of defection is relative to the benefit of cooperation. For the evolution of cooperation, this means that it is more difficult for the spite effect to arise with synergistic payoffs. According to conditions 3a and 3b, natural selection will favor defection whenever  $b$  is high enough. But conditions 9a and 9b state that the spite effect arises only if  $b$  is high enough *and* there are enough cooperators in the population. In the extreme scenario where  $t \rightarrow \infty$  and both  $A(x_C)$  and  $B(x_C)$  are step functions, condition 9b can therefore be satisfied and selection can favor defection only if the number of cooperators falls above the threshold  $k$ . So increasing the benefit of cooperation will favor the spread of defection only if a significant portion of the population is already composed of cooperators.

But assuming that  $A(x_C)$  and  $B(x_C)$  are step functions may be unrealistic in some scenarios. Hence, consider now the behavior of  $A(x_C)$  and  $B(x_C)$  at the other extreme. If  $t \rightarrow 0$ , then  $A(x_C)$  and  $B(x_C)$  take the form of a linear function:

$$A(x_C) = \frac{x_C \cdot a}{N} \quad (11a)$$

$$B(x_C) = \frac{x_C \cdot b}{N} \quad (11b)$$

In this case,  $A(x_C)=a$  and  $B(x_C)=b$  only if  $x_C$  approaches  $N$ . When  $x_C$  approaches  $N$ , conditions **9a** and **9b** are therefore identical to conditions **3a** and **3b**. In the very limited case where  $x_C$  approaches  $N$ , meaning that the population is almost entirely composed of cooperators, increasing the benefit of cooperation does favor the spread of defection.

For other values of  $x_C$ , however, the consequences of increasing the benefit of cooperation are a more delicate matter. To see why, let us first substitute the linear functions  $A(x_C)$  and  $B(x_C)$  in condition **9a** and **9b**. Re-arranging, cooperation evolves if:

$$B(x_C) < A(x_C) \cdot (N - 1) + 1 \tag{12a}$$

And defection evolves if:

$$B(x_C) > A(x_C) \cdot (N - 1) + 1 \tag{12b}$$

Now, replace functions  $A(x_C)$  and  $B(x_C)$  in **12a** and **12b** by the right-hand side of the expressions **11a** and **11b** and simplify. Cooperators will spread provided that:

$$b < a \cdot (N - 1) + \frac{N}{x_C} \tag{13a}$$

And defectors will spread if:

$$b > a \cdot (N - 1) + \frac{N}{x_C} \tag{13b}$$

It is easy to see that the spite effect still persists: higher values of  $b$  increase the selective advantage of defection. So if  $b$  is sufficiently high, then condition **13b** is satisfied and selection favors defection. But conditions **13a** and **13b** place a more severe requirement on the spread of defectors than conditions **3a** and **3b**. When  $x_C$  approaches  $N$ , the term  $N/x_C$  in the right-hand side of Eqs. **13a** and **13b** is equal to 1. In this limiting case, conditions **13a** and **13b** revert back to conditions **3a** and **3b**, as already indicated. But for all other values of  $x_C$ , it is easier for cooperation to evolve. In particular, as  $x_C$  becomes small in comparison to  $N$ , the term  $N/x_C$  becomes large. This makes it easier to satisfy condition **13a** and harder to satisfy condition **13b**, thus favoring the spread of cooperators. When few members of the population cooperate, the spite effect therefore still persists but is not as pronounced as when the population is mostly composed of cooperators.

In sum, synergistic payoffs change how and when the spite effect arises in the prisoner’s delight. Both as  $t \rightarrow \infty$  and as  $t \rightarrow 0$ , the spite effect kicks in with full force only if cooperators already make up a very large portion of the population. If cooperators are rare, then the spite effect weakens as  $t \rightarrow 0$  and is completely absent as  $t \rightarrow \infty$ . Notice also that these values represent extremes that  $t$  can take. Given that changes in the value of  $t$  are associated with continuous changes in the shape of  $B(x_C)$  and  $A(x_C)$ , it is safe to infer that intermediate values of  $t$  yield similar results. That is, the spite effect arises when cooperators abound in the population no matter what value  $t$  takes. But when cooperators are rare, extremely low values of  $t$  dampen the spite effect. As  $t$  goes up, the spite effect is dampened even further until

it eventually vanishes for indefinitely large values of  $t$ . In the finite-population prisoner's delight with synergistic payoffs, increasing the benefit of cooperation therefore does not always favor the spread of defection.

Similar considerations apply to the nearly neutral effect. Substituting  $A(x_C)$  and  $B(x_C)$  for  $a$  and  $b$  in the expression 5 yields:

$$s = \frac{A(x_C) \cdot (N - 1) - B(x_C) + 1}{B(x_C) \cdot (x_C - 1) + x_D} \quad (14)$$

Consider first the case in which the steepness of the function at the inflection point ( $t$ ) goes to  $\infty$ . If  $x_C < k$ , then  $A(x_C) = 0$  and  $B(x_C) = 0$ . In this case, the selection coefficient does not depend on  $b$  or  $a$ ; it only depends on the number of defectors. So the nearly neutral effect is null. If  $x_C \geq k$ , on the other hand, then  $A(x_C) = a$  and  $B(x_C) = b$ . In this case, the selection coefficient does depend on  $b$  and  $a$ ; the nearly neutral effect arises again. But notice that a large number of players must cooperate for the nearly neutral effect to have any bite.

Consider now the case in which  $t$  goes to 0. When  $x_C$  approaches  $N$ ,  $A(x_C) = a$  and  $B(x_C) = b$ . In this case, the selection coefficient in expression 14 approximates the selection coefficient given by expression 5. For these parameter values, high values of  $b$  therefore have roughly the same effect as in the original model: increasing the value of  $b$  reduces the selective pressure for cooperation due to the nearly neutral effect. But this requires that almost all players cooperate. For other values of  $x_C$ , the nearly neutral effect is less pronounced. Although the nearly neutral effect persists as the selection coefficient continues to depend on  $b$ , decreasing  $x_C$  also decreases the value of  $B(x_C)$ . This reduces the extent to which the selection coefficient depends on  $b$ . That is, unless the number of cooperators approaches the total population size, the nearly neutral effect loses some of its force.

It is also worth emphasizing that  $b$  is the benefit of cooperation that goes to the receiver, not the benefit that the donor of cooperation receives from cooperating. In finite populations, each member of the population does not interact with herself. So the composition of the remaining population is affected by the removal of the focal individual. As Forber and Smead observe, this means that in finite populations there is anti-correlation between types: defectors more often interact with cooperators than cooperators do, and cooperators more often interact with defectors than defectors do. As a result, the benefit of cooperation that each cooperator provides tends to go preferentially to other members of the population. Noticing this should make it less surprising that increasing the benefit of cooperation ( $b$ ) favors the spread of defection, especially in small populations. In other words, the spite effect and the nearly neutral effect will be stronger as the population size is smaller and the benefit of cooperation increases. Still, this is an illuminating finding that follows from the model that Forber and Smead analyze.

## Conclusion

When payoffs are linear, increasing the benefit of cooperation hinders the evolution of cooperation in a finite-population version of the prisoner's delight due to the spite effect and the nearly neutral effect. With synergistic payoffs, however, the spite effect and the nearly neutral effect display complex behaviors. If payoffs behave like a stepwise function so that cooperation is beneficial only if a quorum of cooperators is present, then the spite effect and the nearly neutral effect completely vanish when the number of cooperators falls below the quorum threshold. Intuitively, this is because few cooperators cannot provide any benefit at all. So defectors cannot preferentially reap the benefit of cooperation due to the anti-correlation induced by a finite population. If payoff functions have a less extreme shape, on the other hand, then the spite effect and the nearly neutral effect do not disappear entirely. But synergistic payoffs still reduce their strength, especially when cooperators are rare in the population. Again, this is quite intuitive: synergistic payoffs reduce the benefit of cooperation when cooperators are rare, so the benefit of cooperation that defectors preferentially reap is minimal. This is all the more so when there is a very wide range of initial values for the number of cooperators in the population in which the benefit of cooperation is approximately null, and a sharp transition to the regime with a positive benefit.

At this point, it is worth clarifying that none of these observations are guaranteed to hold when payoffs are nonlinear but not synergistic. Take, for example, the proverbial case in which “too many cooks spoil the broth”. In this case, payoffs are a nonlinear function of the number of cooperators in the population but anti-synergistic so that the benefit of cooperation is higher when cooperators are rare. In such a scenario, the spite effect and the nearly neutral effect may be even stronger than in the case of linear payoffs. Although it would be interesting to consider anti-synergistic payoffs in future work, the model that I present here seeks to capture the behavior of cooperative efforts in which more cooperators promote the production of more goods—as in the cases mentioned above of cooperative hunting (Yip et al. 2008; Packer et al. 1990), quorum sensing (Miller and Bassler 2001; Bonner 2009), and some complex human activities (Milinski et al. 2006; Tomasello et al. 2012).

It is also important to emphasize that the model presented here has highly idealized assumptions of its own. For example, it is assumed that the entire population interacts at once. This means that the model cannot represent the case of individuals interacting in groups of different sizes. Clearly, this is an idealized assumption that deviates from what we sometimes expect to observe in nature. It would therefore be interesting to consider how variable group size changes the model results. But a full analytical treatment of models with variable group size can be difficult. So a model with constant group size is a good starting point.

This brings us to a related point. The epistemic value of robust model results has long been a concern in both philosophy and biology (Levins 1966; Orzack and Sober 1993; Weisberg 2006). However, model builders often emphasize some forms of robustness to the detriment of others. In particular, robustness analyses of game-theoretic models commonly focus on what Weisberg and Reisman (2008) call

“parameter robustness” and overlook “structural robustness”. It is therefore standard practice to check whether results hold across a wide range of payoff values. But contrasting forms of interaction and different shapes of the payoff functions are not always considered. There may be good reasons for this: it is often easier and less time-consuming to assess whether results are invariant across a wide range of parameter values (e.g. payoff values) than whether results hold across changes in structural features of the model (e.g. the shape of the payoff function). But to the extent that model results depend on assumptions that deviate from the real-world target that they are supposed to represent, one form of robustness is no more important than the other one.

**Acknowledgements** I thank Alex Rosenberg, Rory Smead, Hannah Read, and especially two anonymous referees for helpful feedback on previous drafts of this paper.

## References

- Allen B, Nowak MA, Wilson EO (2013) Limitations of inclusive fitness. *Proc Natl Acad Sci* 110(50):20135–20139
- Archetti M, Scheuring I (2011) Coexistence of cooperation and defection in public goods games. *Evolution* 65(4):1140–1148
- Bonner JT (2009) *The social amoebae: the biology of cellular slime molds*. Princeton University Press, Princeton
- Cavalli-Sforza LL, Feldman MW (1978) Darwinian selection and “altruism”. *Theor Popul Biol* 14(2):268–280
- Forber P, Smead R (2014) An evolutionary paradox for prosocial behavior. *J Philos* 111(3):151–166
- Forber P, Smead R (2015) Evolution and the classification of social behavior. *Biol Philos* 30(3):405–421
- Hamilton WD (1971) Selection of selfish and altruistic behavior in some extreme models. In: *Man and beast: comparative social behavior*, pp 57–91
- Hardin G (1968) The tragedy of the commons. *Science* 162:1243–1248
- Levins R (1966) The strategy of model building in population biology. *Am Sci* 54(4):421–431
- Maynard-Smith J, Price GR (1973) The logic of animal conflict. *Nature* 246:15–18
- Milinski M, Semmann D, Krambeck H-J, Marotzke J (2006) Stabilizing the Earth’s climate is not a losing game: supporting evidence from public goods experiments. *Proc Natl Acad Sci USA* 103(11):3994–3998
- Miller MB, Bassler BL (2001) Quorum sensing in bacteria. *Annu Rev Microbiol* 55(1):165–199
- Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428(6983):646
- Orzack SH, Sober E (1993) A critical assessment of Levins’s the strategy of model building in population biology (1966). *Q Rev Biol* 68:533–546
- Packer C, Scheel D, Pusey AE (1990) Why lions form groups: food is not enough. *Am Nat* 136(1):1–19
- Smead R, Forber P (2013) The evolutionary dynamics of spite in finite populations. *Evolution* 67(3):698–707
- Taylor C, Fudenberg D, Sasaki A, Nowak MA (2004) Evolutionary game dynamics in finite populations. *Bull Math Biol* 66(6):1621–1644
- Tomasello M, Melis AP, Tennie C, Wyman E, Herrmann E, Gilby IC, Hawkes K, Sterelny K, Wyman E, Tomasello M, Melis A (2012) Two key steps in the evolution of human cooperation: the interdependence hypothesis. *Curr Anthropol* 53(6):673–692
- van Veelen M (2009) Group selection, kin selection, altruism and cooperation: when inclusive fitness is right and when it can be wrong. *J Theor Biol* 259(3):589–600
- Weisberg M (2006) Robustness analysis. *Philos Sci* 73(5):730–742
- Weisberg M, Reisman K (2008) The robust Volterra principle. *Philos Sci* 75(1):106–131

Yip EC, Powers KS, Avilés L (2008) Cooperative capture of large prey solves scaling challenge faced by spider societies. *Proc Natl Acad Sci* 105(33):11818–11822

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