Tragedy of the Commons in the Chemostat

Martin Schuster,[*] Eric Foxall,[†] David Finch,[‡] Hal Smith[§] and Patrick De Leenheer[¶]

May 19, 2017

Abstract

We present a proof of principle for the phenomenon of the tragedy of the commons that is at the center of many theories on the evolution of cooperation. We establish the tragedy in the context of a general chemostat model with two species, the cooperator and the cheater. Both species have the same growth rate function and yield constant, but the cooperator allocates a portion of the nutrient uptake towards the production of a public good -the "Commons" in the Tragedy- which is needed to digest the externally supplied nutrient. The cheater on the other hand does not produce this enzyme, and allocates all nutrient uptake towards its own growth. We prove that when the cheater is present initially, both the cooperator and the cheater will eventually go extinct, hereby confirming the occurrence of the tragedy. We also show that without the cheater, the cooperator can survive indefinitely, provided that at least a low level of public good or processed nutrient is available initially. Our results provide a predictive framework for the analysis of cooperator-cheater dynamics in a powerful model system of experimental evolution.

1 Introduction

Cooperative behaviors abound across all domains of life, from animals to microbes [11, 28]. Yet, their evolution and maintenance is difficult to explain [18, 29, 30]. Why would an individual carry out a costly behavior for the benefit of the group? Cheaters that reap the benefits of cooperation without paying the costs would gain a competitive advantage and invade the population. This conflict of interest between the individual and the group is also known as the tragedy of the commons described by Hardin [15]. To illustrate the tragedy, Hardin considers a scenario first sketched by Lloyd more than 100 years earlier [19], a pasture that is shared by herdsmen. It is in each herdsman's best interest to add additional cattle to the pasture, because

1

he gains the profits from individual cattle sales, but shares the costs of overgrazing with all other herdsmen. This behavior is pursued until, ultimately, the commons is destroyed to the detriment of all.

The problem of cooperation has received considerable attention in the microbial realm [28, 1, 9]. Many microbes perform cooperative behaviors such as biofilm formation, virulence, and collective nutrient acquisition. Often, these behaviors are accomplished by secreted products referred to as public goods [28, 30]. Public goods are costly to produce for the individual but provide a collective benefit to the local group. They include extracellular enzymes that degrade complex food sources, siderophores that scavenge iron from the environment, and secreted toxins and antibiotics that harm other cells. It has been shown in several microbial systems that public goods can be shared within a population of cells, benefitting cells other than the focal producer [6, 14, 22, 24, 13]. For example, when the bacterium *Pseudomonas aeruginosa* is grown on a proteinaceous substrate, mutants deficient in protease secretion enrich in co-culture with the wild-type parent [6, 24]. These non-producing strains are termed obligate cheaters: They cannot grow by themselves, but they have a relative growth advantage in mixed cultures with cooperators. Because cheater enrichment inevitably imposes a burden on the population, the expected outcome is a collapse of the population [23]. This outcome been shown experimentally in a few cases [22, 7, 5]. Often, however, cooperative behaviors are stably maintained and hence, the focus has largely been on mechanisms that avoid a tragedy of the commons [2, 3, 8, 10, 12, 17, 26, 27, 31].

To our knowledge, the notion that obligate cheating behavior constitutes a tragedy of the commons and leads to population collapse has not been mathematically proven. Here, we consider the dynamics between cooperators and obligate cheaters in a continuous culture system. Continuous cultures or chemostats enable microbial culturing at a specified density and growth rate through the constant dilution of the culture with fresh growth medium [25]. There is an extensive mathematical theory that describes population dynamics in the chemostat [25]. We prove that obligate cheaters inevitably increase in frequency until cooperation via public goods is no longer sustainable, eventually leading to wash-out and population collapse. We also show that the dynamics of the cooperators in the absence of cheaters exhibits bistability: Depending on the initial condition of the system, cooperators will either eventually persist, or go extinct. Numerical simulations show that it is possible that the cooperators persist when initially there is no processed nutrient, and only a small level of enzyme. In summary, populations solely comprised of cooperators have a chance to persist, but they are doomed whenever cheaters arise, even at low initial frequency.

2   The model and the tragedy

We propose a chemostat model where S denotes the concentration of the unprocessed nutrient, P of the processed nutrient, E of the enzyme and X_1 is the concentration of the cooperator who produces an enzyme required for nutrient processing, and X_2 of the cheater who does not produce the enzyme. The mass-balance equations for

these variables are as follows:

$$\frac{dS}{dt}(t) = D(t)(S^0(t) - S) - G(E, S) \tag{1}$$

$$\frac{dP}{dt}(t) = G(E, S) - \frac{1}{\gamma}(X_1 + X_2) F(P) - D(t)P \tag{2}$$

$$\frac{dE}{dt}(t) = (1 - q)X_1 F(P) - D(t)E \tag{3}$$

$$\frac{dX_1}{dt}(t) = X_1 (qF(P) - D(t)) \tag{4}$$

$$\frac{dX_2}{dt}(t) = X_2 (F(P) - D(t)) \tag{5}$$

The operating conditions of the chemostat may fluctuate in time, and they are characterized by $D(t)$, the dilution rate, and $S^0(t)$, the concentration of the unprocessed nutrient at the inflow. Both are non-negative functions of time, and additional assumptions for these functions will be introduced below. Unprocessed nutrient is converted into processed nutrient by means of the enzyme. Processed nutrient is produced at rate $G(E, S)$. The per capita consumption rate of processed nutrient by both species is the same, and denoted by $\frac{1}{\gamma}F(P)$, where γ is the yield of this process, taking a value in $(0, 1)$, and which is also assumed to be the same for both species. The cooperator allocates a proportion q, a fixed value in $(0, 1)$, of the processed nutrient it has consumed, towards its own growth. The remaining fraction $(1 - q)$ goes towards the production of the enzyme which is needed to process the unprocessed nutrient. The cheater allocates all processed nutrient it has taken up towards growth. A cartoon of this chemostat model is presented in Figure 1.

We make the following minimal assumptions about the functions G and F:

H1: $G : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is C^1, $G(0, S) = G(E, 0) = 0$ for all $E \geq 0$ and $S \geq 0$, and
$F : \mathbb{R}_+ \to \mathbb{R}_+$ is C^1, and $F(0) = 0$.

This assumption merely implies that there is no conversion of unprocessed nutrient into processed nutrient, when the enzyme or the unprocessed nutrient is missing; similarly there is no growth of either species, or of the enzyme, when the processed nutrient is missing.

For the dilution rate $D(t)$, and input nutrient concentration $S^0(t)$, we assume the following:

H2: The functions $D(t)$ and $S^0(t)$ are continuous for all $t \geq 0$, and there exist positive bounds $\underline{D}$ and $\bar{D}$ such that $\underline{D} \leq D(t) \leq \bar{D}$ for all $t \geq 0$, and positive bounds $\underline{S}^0$ and $\bar{S}^0$ such that $\underline{S}^0 \leq S^0(t) \leq \bar{S}^0$ for all $t \geq 0$.

Our Main Result, which is proved in the Appendix, establishes the tragedy of the commons:

Theorem 1. *Assume that* **H1** *and* **H2** *hold, and assume that the initial condition of* $(1) - (5)$ *is such that* $X_2(0) > 0$; *that is, the cheater is present initially. Then* $(P(t), E(t), X_1(t), X_2(t)) \to (0, 0, 0, 0)$ *as* $t \to \infty$.
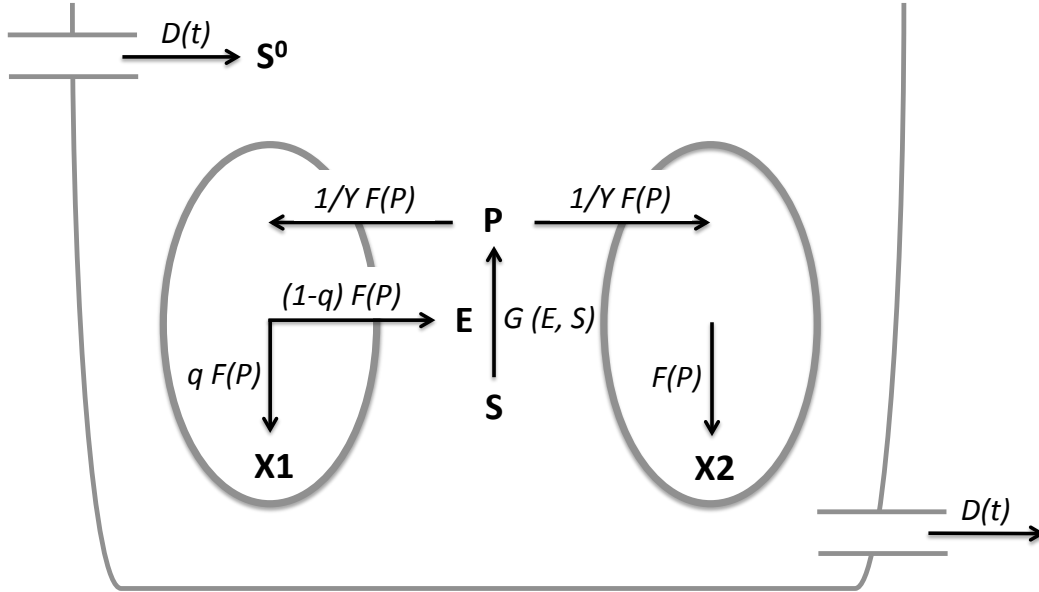
Figure 1: Cartoon of the chemostat with two competing bacterial types: Species are indicated in bold and rates are indicated in italics. X_1, cooperator; X_2, cheater; S, nutrient substrate; S^0, unprocessed nutrient substrate in inflow; P, processed nutrient; E, enzyme; $D(t)$, dilution rate, $1/yF(P)$, per capita nutrient consumption rate; $F(P)$, growth rate; q and $1 - q$, proportions of nutrient allocated towards growth and enzyme production, respectively.

Figure 2 depicts the tragedy in case of mass action kinetics $G(E, S) = kES$, and Monod uptake function $F(P) = mP/(a + P)$. The equations have been scaled such that S^0 and D are both constant equal to one. Initial data are as follows: $S(0) = 1, P(0) = 0, E(0) = 0.8, X_1(0) = 0.2, X_2(0) = 0.03$. The cooperator peaks early and declines sharply as the cheater continues to thrive, reaching a maximum followed by a rapid decline.

We show next that the tragedy also occurs in cases where the processing of the substrate into processed nutrient proceeds in more than one step. First, let us single out the biochemical reaction taking place in model $(1) - (5)$. Borrowing notation from (bio)chemistry, this reaction can be represented as follows:

$$S + E \rightarrow P + E,$$
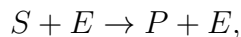
where the reaction rate of formation of processed nutrient is $g(e, s)$, expressed in rescaled variables (see the Appendix for the rescaling). If we would only model this process, and ignore enzyme production, inflow of substrate, and outflow of substrate,
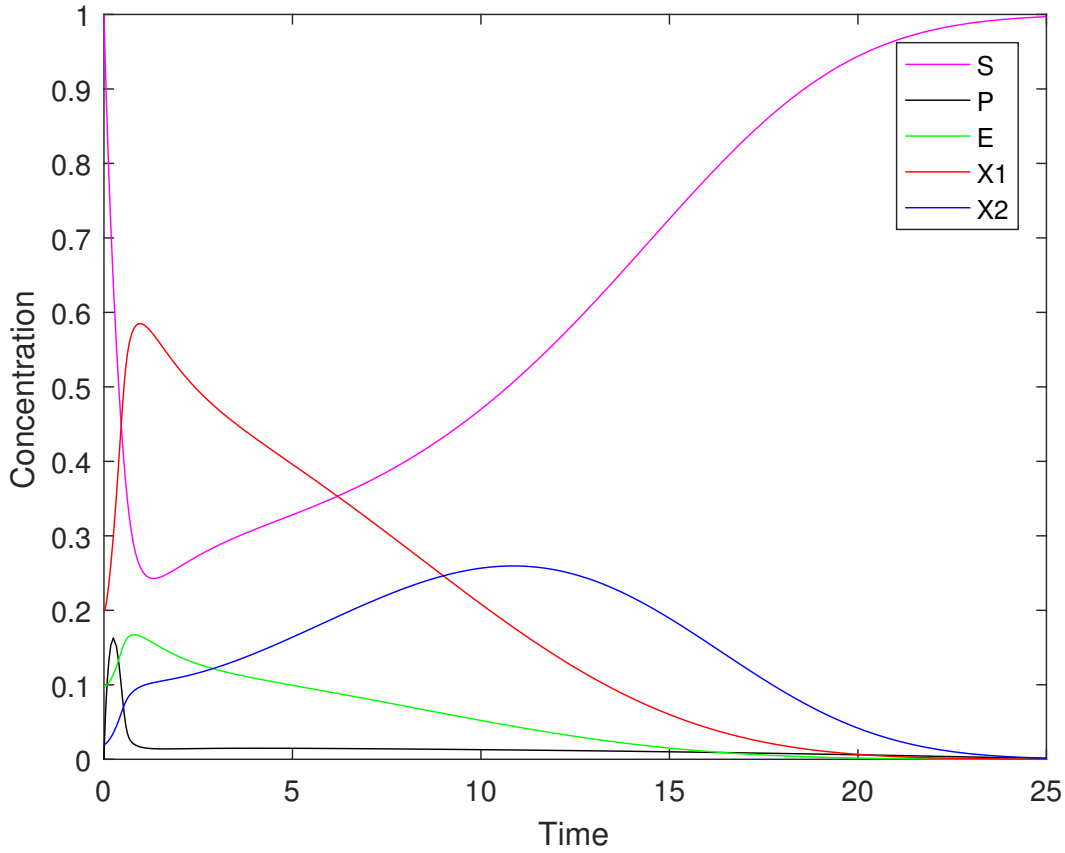
Figure 2: Time series of the components of system $(1) - (5)$, where $S^0 = 1$, $D = 1$, $q = 0.8$, $\gamma = 1$, $G(E, S) = kES$, $F(P) = mP/(a + P)$ with $k = 20$, $m = 5$ and $a = 0.05$. Initial data: $S(0) = 1, P(0) = 0, E(0) = 0.1, X_1(0) = 0.2, X_2(0) = 0.02$.

enzyme and processed nutrient, we would have the following mass balance:

$$
\begin{aligned}
\frac{ds}{dt}(t) &= -g(e, s) \\
\frac{de}{dt}(t) &= 0 \\
\frac{dp}{dt}(t) &= g(e, s)
\end{aligned}
$$

Suppose now that the biochemistry describing the conversion of substrate into processed nutrient takes occurs via an intermediate step:

$$S + E \longleftrightarrow C \to P + E,$$
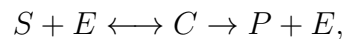
where C represents an intermediate complex formed by the action of the enzyme on the substrate. Let us for simplicity assume that the reaction rates are of the mass action type (with respective rate constants k_1 and k_{-1} for the first reversible reaction, and k_2 for the second reaction), then the mass balance model for this 2-step

biochemical reaction network is:

$$\frac{ds}{dt}(t) = -k_1 es + k_{-1}c$$

$$\frac{de}{dt}(t) = -k_1 es + k_{-1}c + k_2 c$$

$$\frac{dc}{dt}(t) = k_1 es - k_{-1}c - k_2 c$$

$$\frac{dp}{dt}(t) = k_2 c$$

The key property for this network is the conservation of the following quantity:

$$s(t) + e(t) + 2c(t) + p(t),$$

which is easily verified by showing that its derivative with respect to time is zero. If we integrate this biochemical reaction network in our chemostat model, then we obtain the following scaled chemostat model:

$$\frac{ds}{dt}(t) = D(t)(S^0(t) - s) - k_1 es + k_{-1}c \tag{6}$$

$$\frac{dp}{dt}(t) = k_2 c - (x_1 + x_2)\, f(p) - D(t)p \tag{7}$$

$$\frac{de}{dt}(t) = (1 - q)x_1 f(p) - k_1 es + k_{-1}c + k_2 c - D(t)e \tag{8}$$

$$\frac{dc}{dt}(t) = k_1 es - k_{-1}c - k_2 c - D(t)c \tag{9}$$

$$\frac{dx_1}{dt}(t) = x_1 \left(q f(p) - D(t) \right) \tag{10}$$

$$\frac{dx_2}{dt}(t) = x_2 \left(f(p) - D(t) \right) \tag{11}$$

We show in the last section of the Appendix that the tragedy continues to hold, in the sense that the conclusion of Theorem 1 remains valid for this more general system.

Of course, more complicated biochemical reaction networks of the digestion process, with multiple intermediate complexes $C_1, \ldots C_n$:

$$S + E \longleftrightarrow C_1 \longleftrightarrow \ldots \longleftrightarrow C_n \to P + E$$
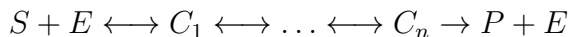
could be used here instead, and the tragedy would continue to hold in such cases. The key property is that the mass balance equations corresponding to these networks should exhibit a conservation law to guarantee the boundedness of the solutions of the chemostat model which integrates this biochemistry. Most reasonable biochemical reaction networks do indeed possess such conservation laws.

3   Cooperators can persist when cheaters are absent

We have shown that when cheaters are present initially, the total population of cooperators and cheaters, is doomed. Next we investigate what happens when cheaters

are absent by considering a special case of the chemostat model $(1) - (5)$ with $X_2 = 0$, and constant operating parameters D and S^0, which are both assumed to be positive:

$$\frac{dS}{dt}(t) = D(S^0 - S) - EG(S) \tag{12}$$

$$\frac{dP}{dt}(t) = EG(S) - \frac{1}{\gamma}X_1 F(P) - DP \tag{13}$$

$$\frac{dE}{dt}(t) = (1 - q)X_1 F(P) - DE \tag{14}$$

$$\frac{dX_1}{dt}(t) = X_1 (qF(P) - D) \tag{15}$$

Notice that the nutrient processing rate has been specialized to $EG(S)$, implying that it is proportional to the enzyme concentration E, and a possibly nonlinear function of the nutrient $G(S)$. We replace assumption **H1**, by the following assumption, which introduces a monotonicity condition for F, and monotonicity and concavity condition for G:

H1': $G : \mathbb{R}_+ \to \mathbb{R}_+$ is C^2, $G(0) = 0$, $dG/dS(S) > 0$ for all $S > 0$, and
$d^2G/dS^2(S) \leq 0$ for all $S \geq 0$, and
$F : \mathbb{R}_+ \to \mathbb{R}_+$ is C^1, $F(0) = 0$, $dF/dP(P) > 0$ for all $P > 0$.

The concavity condition for G will be used to limit the number of steady states of this system. The most commonly used choices for the functions for F and G are Monod functions (i.e. $F(P) = mP/(a+P)$, where a and m are positive parameters), which satisfy these assumptions. But note that a linear function $G(S) = kS$, with $k > 0$ is allowed as well. In other words, the processing rate of nutrient (per unit of enzyme) does not necessarily have to saturate for large S-values.

The following dichotomy -global extinction, or bistability- is proved in the Appendix, and shows that the cooperator may persist when there are no cheaters; it refers to a scalar, nonlinear equation (43), which is given in the Appendix as well.

Theorem 2. *Suppose that* **H1'** *holds, and that* $P^* := F^{-1}\left(\frac{D}{q}\right) < S^0$.

1. *If equation (43) has no solutions, then the washout steady state $(0, 0, 0, 0)$ is globally asymptotically stable for system $(12) - (15)$.*

2. *If equation (43) has two distinct solutions, then system $(12) - (14)$ has 3 steady states, the washout steady state $(0, 0, 0, 0)$ and two positive steady states E_1 and E_2. The washout steady state and E_2 are locally asymptotically stable, and E_1 is a saddle with a three-dimensional stable manifold, and one-dimensional unstable manifold. The stable manifold is the common boundary of the regions of attraction of the washout steady state and E_2. Every solution of system $(12) - (15)$ converges to one of the three steady states. Persistence of the cooperator occurs for all initial conditions contained in the region of attraction of E_2, and initial conditions on the stable manifold of the saddle E_1.*

7

Figure 3 illustrates the persistence of the cooperator in the absence of cheaters, even when there is no processed nutrient, and only a little amount of enzyme initially. Notice that the initial condition used in the simulation for Figure 3 is the same as the initial condition used for Figure 2, and the model parameters are the same as well. Nevertheless, the fate of the cooperator is very different: it goes extinct when the cheater is present initially (Figure 2), but persists otherwise (Figure 3).

4 Conclusion

Although the tragedy of the commons is such a pervasive notion in the recent developments of theories about the evolution of cooperation, we were unable to find any mathematical models that have rigorously analyzed an important group-level effect: the collapse of a population as a consequence of the dynamic interaction between cooperating and cheating individuals. Here we have proved mathematically that the tragedy of the commons occurs in a chemostat system with cooperators that supply a public good required for growth, and cheaters that do not. The sole difference between cooperators and cheaters in this system is the cost associated with public good production, which is only experienced by the cooperator. While the cooperator diverts a fraction of the ingested nutrient from growth to public good production, the cheater invests everything in growth. We assume that there are no pleiotropic costs to cheating, and that the environment is well mixed, disregarding spatial structure as a major factor that promotes cooperation [29, 30]. Our results support the occurrence of the tragedy of the commons as a consequence of the selfish actions of individuals that result in the complete collapse of the shared public good [15, 23]. When this public good is essential for growth, the tragedy is manifested by the extinction of the whole group [22, 8, 5].

To understand how the tragedy of the commons arises in the chemostat, we perform a simple thought-experiment. Assume that initially there are no cheaters $(X_2(0) = 0)$, and suppose that the assumptions of Theorem 2 hold. If the initial condition of system $(12) - (15)$ is contained in the region of attraction of the locally stable steady state E_2, the solution will converge to, and eventually settles at this steady state. Numerical simulations (Figure 3) show that this can happen even if there is only a low initial amount of enzyme ($E(0)$ is small), and no initial processed nutrient ($P(0) = 0$). The cooperator-only population therefore persists. However, if cheaters do suddenly appear -for example by mutation or by invasion into the environment- even in extremely low numbers, Theorem 1 shows that the total population of cooperators and cheaters is doomed, confirming the tragedy of the commons. One of the two proofs of Theorem 1 gives clues on how this happens: The *ratio of cooperators to cheaters* will always decrease. It may appear as if the cheaters will overtake the cooperators, and at least for a while, this is indeed what happens. However, in the long run there are not enough cooperators around to produce the enzyme levels required for nutrient processing, and this leads to the extinction of cheaters and cooperators alike.

To put our results in context, it is useful to quote from Hardin's original interpretation of the tragedy, see [15]:

The tragedy of the commons develops in this way. Picture a pasture open to all.
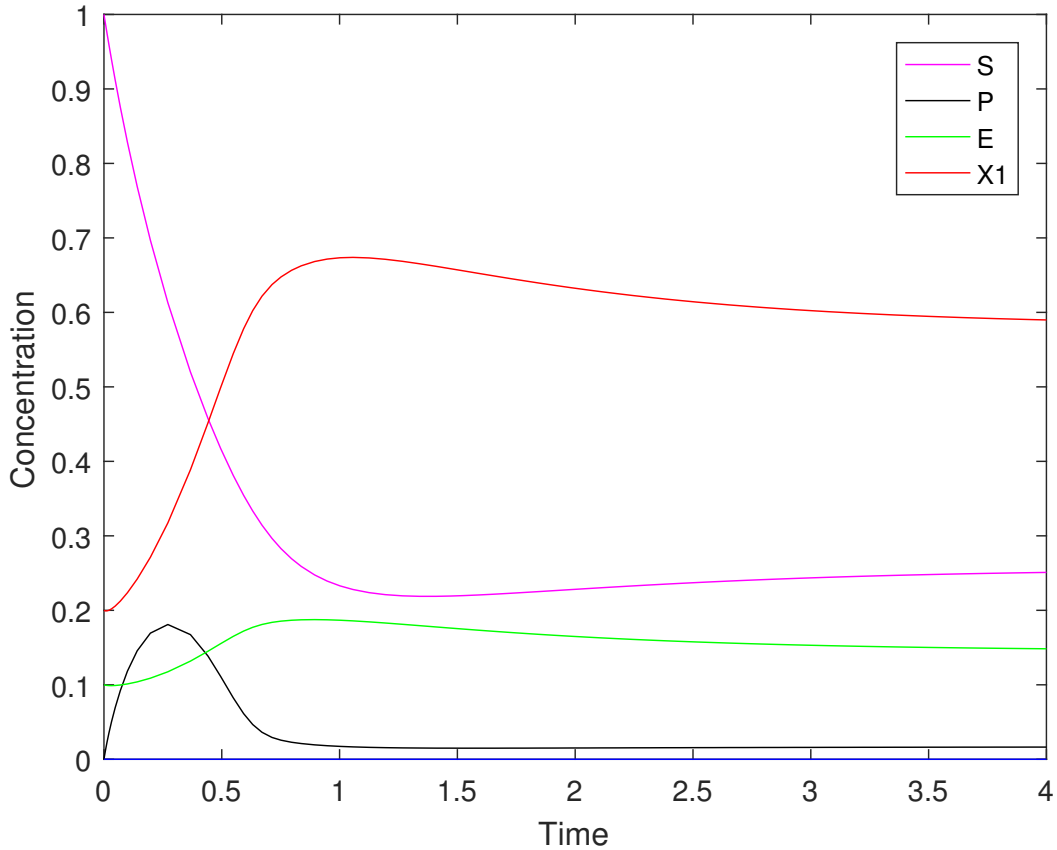
Figure 3: Time series of the components of system $(12) - (15)$, where $S^0 = 1$, $D = 1$, $q = 0.8$, $\gamma = 1$, $G(E, S) = kES$, $F(P) = mP/(a + P)$ with $k = 20$, $m = 5$ and $a = 0.05$. Initial data: $S(0) = 1, P(0) = 0, E(0) = 0.1, X_1(0) = 0.2$.

It is to be expected that each herdsman will try to keep as many cattle as possible on the commons. Such an arrangement may work reasonably satisfactorily for centuries because tribal wars, poaching, and disease keep the numbers of both man and beast well below the carrying capacity of the land. Finally, however, comes the day of reckoning, that is, the day when the long-desired goal of social stability becomes a reality. At this point, the inherent logic of the commons remorselessly generates tragedy.

As a rational being, each herdsman seeks to maximize his gain. Explicitly or implicitly, more or less consciously, he asks, "What is the utility to me of adding one more animal to my herd?" This utility has one negative and one positive component:

1. *The positive component is a function of the increment of one animal. Since the herdsman receives all the proceeds from the sale of the additional animal, the positive utility is nearly +1.*

2. *The negative component is a function of the additional overgrazing created by one or more animal. Since, however, the effects of overgrazing are shared by all herdsman, the negative utility for any particular decision making herdsman*

is only a fraction of -1.

Adding together the component partial utilities, the rational herdsman concludes that the only sensible course for him to pursue is to add another animal to his herd. And another; and another.... But this is the conclusion reached by each and every rational herdsman sharing a commons. Therein is the tragedy. Each man is locked into a system that compels him to increase his herd without limit-in a world that is limited. Ruin is the destination toward which all men rush, each pursuing his own best interest in a society that believes in the freedom of the commons. Freedom in a commons brings ruin to all.

It is interesting to note that Hardin's verbal description of the tragedy makes no explicit distinction between cooperators and cheaters, which is in contrast with recent interpretations of the tragedy in evolutionary biology [23]. In natural populations there are many different ways individuals can cooperate or cheat, and clearly articulating the distinction between both types is necessary to correlate it to the occurrence of the tragedy [23]. In its essence, the tragedy of the commons is the depletion of a common resource or a public good by the selfish action of competing individuals, thereby decreasing the average fitness of all individuals.

In the realm of game theory, the tragedy is described by a public good game or an N-person prisoner's dilemma [16]. In these types of games, selfishness is the superior strategy or Nash equilibrium [16]. While it can predict winning strategies, however, game theory does not generally consider the feedback of individual behavior phenotypes on group productivity.

According to [23], the exploitation of different types of resources can give rise to a tragedy of the commons. The first, which fits Hardin's analogy described above, involves the selfish exploitation of a common, extrinsic resource to the point of complete depletion, which causes all individuals to perish. The second type involves resources that are themselves the product of social behavior. In this case, the resource is a public good that is either formed by cooperation, or by restraining from conflict. Cooperation via public goods is pervasive in microbial social behavior, and it is also the case that we have described here with our model $(1) - (5)$. As we have seen, the tragedy arises when non-cooperating cheaters reap the benefits provided by cooperators, without paying the costs. Microbial cooperative behaviors vulnerable to cheating include extracellular secretions such as enzymes and metabolites [12, 20, 22]. A particularly compelling example is the altruistic investment in the non-spore parts of a multicellular fruiting body in myxobacteria [7].

A different, more abstract, type of public good involves individuals restraining from potential conflict. A tragedy arises if the costs invested in compettitive behavior decrease overall productivity. In this case, less emphasis is placed on the depletion of extrinsic resources. A relevant example comes from another chemostat study which investigated the outcome of social conflict between different metabolic strategies in yeast, respiration and fermentation [20]. Respirers use glucose slowly but efficiently, whereas fermenters use glucose fast but wastefully. Thus, respiration is the strategy that provides the highest group-level benefit. Nevertheless, as shown experimentally and confirmed by simulation, fermenters are favored and fully displace respirers during glucose-limited growth in a chemostat [20]. Notably, in this system, as in restraint from conflict in general, one strategy does not obligately

10

depend on the other for its success.

As we have proven in this study, population collapse is inevitable in an obligate relationship, because the cooperator to cheater ratio always decreases. Eventually the cheater becomes so dominant that too little public good is produced by the cooperator, leading to the extinction of both types. The differential equation framework presented here will permit the in-depth analysis of mechanisms that promote cooperation. The contribution of specific parameters or functional forms can be investigated. For example, how much higher would growth yield or nutrient uptake rates have to be in a cooperator compared to a cheater to make public good cooperation sustainable?

Appendix

Proof of Theorem 1

By scaling the state variables of system $(1) - (5)$ as follows:

$$
\begin{aligned}
s &= S \\
p &= P \\
e &= \frac{E}{\gamma} \\
x_1 &= \frac{X_1}{\gamma} \\
x_2 &= \frac{X_2}{\gamma},
\end{aligned}
$$

and introducing the rescaled functions

$$
\begin{aligned}
g(e, s) &:= G(\gamma e, s) \\
f(p) &:= F(P),
\end{aligned}
$$

we obtain the following scaled model:

$$
\frac{ds}{dt}(t) = D(t)(S^0(t) - s) - g(e, s) \tag{16}
$$

$$
\frac{dp}{dt}(t) = g(e, s) - (x_1 + x_2)f(p) - D(t)p \tag{17}
$$

$$
\frac{de}{dt}(t) = (1 - q)x_1 f(p) - D(t)e \tag{18}
$$

$$
\frac{dx_1}{dt}(t) = x_1 (qf(p) - D(t)) \tag{19}
$$

$$
\frac{dx_2}{dt}(t) = x_2 (f(p) - D(t)) \tag{20}
$$

Notice that **H1**, which holds for the rate functions $G(E, S)$ and $F(P)$, is also valid for the scaled rate functions $g(e, s)$ and $f(p)$.

The total mass of this scaled model,

$$
m = s + p + e + x_1 + x_2,
$$

satisfies a linear equation:

$$
\frac{dm}{dt}(t) = D(t)(S^0(t) - m), \tag{21}
$$

which is easily verified by adding all the equations of the scaled model. This equation, and the upper bound for $S^0(t)$ in **H2** imply that the following family of compact sets

$$
\Omega_\epsilon = \{(s, p, e, x_1, x_2) \mid s \geq 0, p \geq 0, e \geq 0, x_1 \geq 0, x_2 \geq 0, m \leq \bar{S}^0 + \epsilon\},
$$

are forward invariant sets of the scaled model, for all $\epsilon \geq 0$.

The Main Result, Theorem 1, is an immediate Corollary of the following result, which is the tragedy of the commons for the scaled model:

12

Theorem 3. *Assume that* **H1** *and* **H2** *hold, and assume that the initial condition of* (16) − (20) *is such that* $x_2(0) > 0$; *that is, the cheater is present initially. Then* $(p(t), e(t), x_1(t), x_2(t)) \rightarrow (0, 0, 0, 0)$ *as* $t \rightarrow \infty$.

Proof. Given the initial condition, we can find an $\epsilon \geq 0$ such that the solution $(s(t), p(t), e(t), x_1(t), x_2(t))$ is contained in the compact set Ω_ϵ for all $t \geq 0$. We shall present two proofs. The first involves a (biologically nontrivial) transformation of one of the system's variables. The second considers the ratio of cooperators and cheaters, a biologically natural measure, and reveals that this ratio does not increase.

Proof 1: Consider the variable $y_2 = x_2^q$. Then

$$\frac{dy_2}{dt}(t) = y_2(qf(p) - qD(t))$$

Equation (19), and the above equation can be integrated:

$$
\begin{aligned}
x_1(t) &= x_1(0) \, e^{\int_0^t qf(p(\tau)) - D(\tau) d\tau} \\
y_2(t) &= y_2(0) \, e^{\int_0^t qf(p(\tau)) - qD(\tau) d\tau} > 0, \text{ for all } t \text{ since } y_2(0) = x_2^q(0) > 0,
\end{aligned}
$$

Dividing the first by the second equation yields:

$$x_1(t) = y_2(t) \frac{x_1(0)}{y_2(0)} e^{-(1-q) \int_0^t D(\tau) d\tau} \leq B \frac{x_1(0)}{y_2(0)} e^{-(1-q)\underline{D}t},$$

where we have used the lower bound for $D(t)$, see **H2**, to establish the last inequality, and the positive bound B for $y_2(t)$ which exists because the solution, and therefore also $x_2(t)$, is bounded. From this follows that $\lim_{t \to \infty} x_1(t) = 0$, where the convergence is at least exponential with rate $(1 - q)\underline{D}$.

Next we consider the dynamics of the variable $z = Qx_1 - e$, where $Q = (1 - q)/q$:

$$\dot{z} = -D(t)z,$$

which is solvable, yielding $z(t) = z(0) \, e^{-\int_0^t D(\tau) d\tau}$. The lower bound $\underline{D}$ for $D(t)$ in **H2**, then implies that $z(t) \rightarrow 0$ at a rate which is at least exponential with rate $\underline{D}$. This fact, together with the convergence of $x_1(t)$ to zero established above, implies that $e(t) \rightarrow 0$ as well.

Next, consider the p-equation (17). There holds that for each $\tilde{\epsilon} > 0$:

$$\frac{dp}{dt}(t) \leq \tilde{\epsilon} - \underline{D}p, \text{ for all sufficiently large } t.$$

Notice that we used that $g(0, s) = 0$ for all $s \geq 0$, and the continuity of g, see **H1**, as well as **H2** for the lower bound of $D(t)$. It follows that $\limsup_{t \to \infty} p(t) \leq \tilde{\epsilon}/\underline{D}$, and since $\tilde{\epsilon} > 0$ was arbitrary, there follows that $p(t) \rightarrow 0$.

Finally, we consider the x_2-equation (20). Since $p(t) \rightarrow 0$ and $f(0) = 0$ by **H1**, there holds that $f(p(t)) \leq \underline{D}/2$ for all t sufficiently large. Consequently,

$$\frac{dx_2}{dt}(t) \leq -\frac{\underline{D}}{2}x_2, \text{ for all sufficiently large } t,$$

13

and thus $x_2(t) \to 0$, concluding the proof in this case.

Proof 2: Equations (19) and (20) can be integrated:

$$x_1(t) = x_1(0)\, e^{\int_0^t qf(p(\tau))-D(\tau)d\tau} \tag{22}$$

$$x_2(t) = x_2(0)\, e^{\int_0^t f(p(\tau))-D(\tau)d\tau} > 0, \text{ for all } t \text{ since } x_2(0) > 0. \tag{23}$$

Thus, the ratio $r(t) = x_1(t)/x_2(t)$ is well-defined and satisfies the differential equation:

$$\frac{dr}{dt}(t) = -(1-q)f(p)r,$$

which shows that the ratio does not increase. The solution of this equation is:

$$r(t) = r(0)\, e^{-(1-q)\int_0^t f(p(\tau))d\tau} \tag{24}$$

We distinguish two cases depending on the integrability of the function $f(p(t))$:

Case 1: $\int_0^\infty f(p(\tau))d\tau = \infty$.
It follows from (24) that $r(t) \to 0$, and hence also $x_1(t) \to 0$ because $x_2(t)$ is bounded. Proof of convergence of $e(t), p(t)$ and $x_2(t)$ to zero now proceeds as in **Proof 1**.

Case 2: $\int_0^\infty f(p(\tau))d\tau < \infty$.
It follows from (22)−(23) that both $x_1(t) \to 0$ and $x_2(t) \to 0$, because $0 < \underline{D} \le D(t)$ for all t, by **H2**. Proof of convergence of $e(t)$ and $p(t)$ to zero now proceeds as in **Proof 1** as well.

$\square$

Proof of Theorem 2

By scaling the state variables of system (12) − (15) in the usual way as follows:

$$
\begin{aligned}
s &= S \\
p &= P \\
e &= \frac{E}{\gamma} \\
x_1 &= \frac{X_1}{\gamma},
\end{aligned}
$$

and switching to lower case letters for the rate functions:

$$
\begin{aligned}
g(s) &:= G(S) \\
f(p) &:= F(P),
\end{aligned}
$$

and for the chemostat's constant operating parameters:

$$
\begin{aligned}
d &:= D \\
s^0 &:= S^0,
\end{aligned}
$$

14

we obtain the following scaled model:

$$\frac{ds}{dt}(t) = d(s^0 - s) - eg(s) \tag{25}$$

$$\frac{dp}{dt}(t) = eg(s) - x_1 f(p) - dp \tag{26}$$

$$\frac{de}{dt}(t) = (1 - q)x_1 f(p) - de \tag{27}$$

$$\frac{dx_1}{dt}(t) = x_1 (qf(p) - d) \tag{28}$$

Notice that **H1'**, which holds for the rate functions $G(S)$ and $F(P)$, is also valid for the rate functions $g(s)$ and $f(p)$.

We introduce two new variables:

$$m = s + p + e + x_1 \tag{29}$$

$$v = e - Qx_1, \text{ where } Q := \frac{1 - q}{q} \tag{30}$$

and choose to drop the s and e-equations from system $(25) - (28)$, transforming it to:

$$\frac{dm}{dt}(t) = d(s^0 - m) \tag{31}$$

$$\frac{dv}{dt}(t) = -dv \tag{32}$$

$$\frac{dp}{dt}(t) = (v + Qx_1)g(m - p - v - x_1/q) - xf(p) - dp \tag{33}$$

$$\frac{dx_1}{dt}(t) = x_1 (qf(p) - d) \tag{34}$$

with state space $\{p, x_1 \geq 0 : m \geq p + v + x_1/q, \ v + Qx_1 \geq 0\}$, which is forward invariant. The variables $m(t)$ and $v(t)$ converge exponentially to s^0 and 0 respectively, hence it is natural to study the limiting system:

$$\frac{dp}{dt}(t) = Qx_1 g(s^0 - p - x_1/q) - x_1 f(p) - dp \tag{35}$$

$$\frac{dx_1}{dt}(t) = x_1(qf(p) - d) \tag{36}$$

which is defined on the state space $\{p, x_1 \geq 0 : p + x_1/q \leq s^0\}$, which is forward invariant. It turns out to be more convenient to transform this system using the variable:

$$w = p + \frac{1}{q}x_1, \tag{37}$$

instead of the variable p, yielding:

$$\frac{dw}{dt}(t) = Qx_1 g(s^0 - w) - dw \tag{38}$$

$$\frac{dx_1}{dt}(t) = x_1(qf(w - x_1/q) - d) \tag{39}$$

15

with state space $\Omega_{red} = \{x_1 \geq 0 : x_1/q \leq w \leq s^0\}$, which is forward invariant.

We start our analysis of system $(38) - (39)$ by determining the nullclines. The w-nullcline is given by:

$$x_1 = h(w), \text{ where } h(w) = \frac{d}{Q} \frac{w}{g(s^0 - w)}. \tag{40}$$

The main properties of the function $h(w) : [0, s^0) \to \mathbb{R}_+$ are:

1. $h(0) = 0$, and $\lim_{w \to s^0} h(w) = \infty$.

2. $h'(w) = \frac{d}{Q} \frac{g(s^0 - w) + wg'(s^0 - w)}{g^2(s^0 - w)} > 0$, by **H1'**.

3. $h''(w) = \frac{d}{Q} \frac{-wg''(s^0 - w)g^2(s^0 - w) + 2g(s^0 - w)g'(s^0 - w)\left(g(s^0 - w) + wg'(s^0 - w)\right)}{g^4(s^0 - w)} > 0$, by **H1'**.

Thus, the function $h(w)$ is zero at zero, is increasing with a vertical asymptote at $w = s^0$, and it is strictly convex.

To obtain a nontrivial x_1-nullcline in the state space, we make one more assumption, namely:

$$p^* := f^{-1}\left(\frac{d}{q}\right) \text{ satisfies } p^* < s^0. \tag{41}$$

This assumption merely expresses that the cooperator has a break-even steady state concentration for the processed nutrient at a level below the input nutrient concentration s^0, see equation (28). In addition to the horizontal axis $x_1 = 0$, there is a nontrivial x_1-nullcline which is particularly easy to express using p^*, as the graph of a linear function:

$$x_1 = q(w - p^*). \tag{42}$$

Any nonzero steady states of the limiting system are given by the intersection of the w- and the nontrivial x_1-nullcine, which are determined by the solutions of the equation:

$$h(w) = q(w - p^*), \ 0 \leq w < s^0. \tag{43}$$

In view of the convexity of the function h, there are either no, one, or two solutions to (43), and generically there are none, or two. We will construct the phase portrait of the limiting system in these two cases.

Lemma 1. *Suppose that* **H1'** *and* (41) *hold.*

1. *If equation* (43) *has no solutions, then system* $(38) - (39)$ *has a unique steady state* $(0, 0)$ *which is globally asymptotically stable with respect to initial conditions in* Ω_{red}.

2. *If equation* (43) *has two solutions* $w_1 < w_2$, *then system* $(38) - (39)$ *has 3 steady states,* $(0, 0)$, $(w_1, h(w_1))$ *and* $(w_2, h(w_2))$. *The steady states* $(0, 0)$ *and* $(w_2, h(w_2))$ *are locally asymptotically stable, and* $(w_1, h(w_1))$ *is a saddle with one-dimensional stable manifold* W_s, *and one-dimensional unstable manifold* W_u. *The stable manifold* W_s *intersects the boundary of* Ω_{red} *in two points, one on the boundary* $x_1 = qw$, *the other on the boundary* $w = s^0$, *forming a separatrix: Initial conditions below* W_s *give rise to solutions converging to* $(0, 0)$, *whereas initial conditions above* W_s *give rise to solutions converging to* $(w_2, h(w_2))$, *yielding bistability in the limiting system, see Figure 4.*
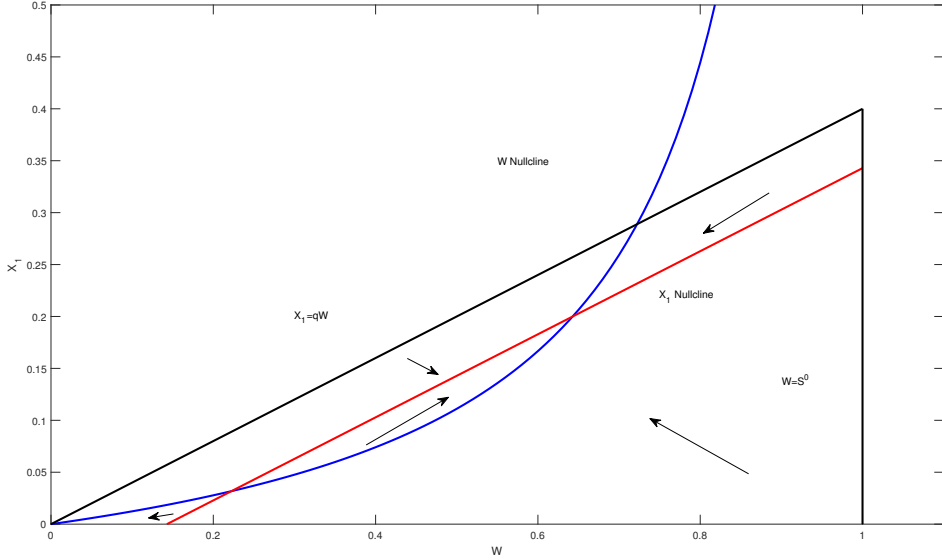
16

Figure 4: Phase plane of the limiting system $(38)-(39)$ in case there are two nonzero steady states.

Proof.   1. If (43) has no solutions, then the w-nullcline and the nontrivial x_1-nullcline do not intersect, and thus $(0,0)$ is the only steady state of the system. The state space Ω_{red} is divided in 3 parts by the two nullclines, and it is easy to see that the region enclosed between both nullclines and the boundary of Ω_{red} is a trapping region in which solutions monotonically converge to the zero steady state. Solutions starting in the region above the w-nullcline are monotonically decreasing (increasing) in the x_1-component (w-component), but since that region does not contain nontrivial steady states, these solutions must enter the trapping region between both nullclines. Similarly, solutions that start below the x_1-nullcline are monotonically increasing (decreasing) in the x_1-component (w-component), and must enter the trapping region as well. This concludes the proof of the assertion that $(0,0)$ is a globally asymptotically stable steady state.

2. If equation (43) has two solutions $w_1 < w_2$, then the nullclines intersect in two distinct points, yielding the positive steady states $(w_1, h(w_1))$ and $(w_2, h(w_2))$, see Figure 4. The third steady state is $(0,0)$. It is not hard to see that the state space is now divided into 5 parts, 3 of which are trapping regions. Each of these trapping regions is enclosed by arcs of the nullclines or segments of the boundary of Ω_{red} which either connect pairs of steady states, or a steady state and a point on the boundary of the state space Ω_{red}, see Figure 4. There are also 2 remaining regions, which we call the NW and SE regions, for obvious reasons.

To complete the phase plane analysis, we perform a linearization of the system

17

at the steady states. The Jacobian matrix of the limiting system is

$$\begin{pmatrix} -Qx_1g'(s^0-w)-d & Qg(s^0-w) \\ x_1qf'(w-x_1/q) & (qf(w-x_1/q)-d)-x_1f'(w-x_1/q) \end{pmatrix}$$

We focus on the middle steady state $(w_1, h(w_1))$, where the Jacobian evaluates to:

$$J_1 = \begin{pmatrix} -d\left(\frac{w_1g'(s^0-w_1)}{g(s^0-w_1)}+1\right) & Qg(s^0-w) \\ \frac{d}{Q}q\frac{w_1}{g(s^0-w_1)}f'(p^*) & -\frac{d}{Q}\frac{w_1}{g(s^0-w_1)}f'(p^*) \end{pmatrix}$$

Clearly, the trace is negative, and the determinant is given by:

$$\frac{d}{Q}\frac{w_1}{g(s^0-w_1)}f'(p^*)\left[d\left(\frac{w_1g'(s^0-w_1)}{g(s^0-w_1)}+1\right)-qQg(s^0-w_1)\right]$$

We claim that this determinant is negative, which implies that this steady state is a saddle. This follows from the fact that the slope of the tangent line to the graph of the convex function $h(w)$ at $w = w_1$ must be smaller than the slope of the line $x_1 = q(w - p^*)$, which is of course q:

$$h'(w_1) < q.$$

Recalling the derivative of $h(w)$ given above, it can be shown that this latter inequality is equivalent to the expression in the square brackets in the determinant being negative, which establishes the claim.

Incidentally, a similar argument can be used to show that the determinant of the linearization at the steady state $(w_2, h(w_2))$ is positive, because in this case $h'(w_2) > q$. Since the trace of that linearization is also negative, this shows that $(w_2, h(w_2))$ is locally asymptotically stable. The linearization at $(0,0)$ is triangular with both diagonal entries equal to $-d$, from which also follows that $(0,0)$ is locally asymptotically stable.

Now we turn to the question of the location of the one-dimensional stable and unstable manifolds W_s, respectively W_u, of the saddle. Therefore, we determine the eigenvectors of the negative eigenvalue λ_1, and the positive eigenvalue λ_2 of the Jacobian matrix J_1. We have that

$$J_1\begin{pmatrix} 1 \\ r_1 \end{pmatrix} = \lambda_1\begin{pmatrix} 1 \\ r_1 \end{pmatrix} \text{ with } \lambda_1 < 0, \text{ and } J_1\begin{pmatrix} 1 \\ r_2 \end{pmatrix} = \lambda_2\begin{pmatrix} 1 \\ r_2 \end{pmatrix}, \text{ with } \lambda_2 > 0,$$

where we wish to determine, or at least estimate, r_1 and r_2. We can find r_1 by considering the first of the two equations determining λ_1:

$$r_1 = \frac{1}{g(s^0-w)}\left(\lambda_1 + d\left(\frac{w_1g'(s^0-w_1)}{g(s^0-w_1)}+1\right)\right).$$

We claim that $r_1 < 0$. Indeed, since the trace of J_1 (which equals $\lambda_1 + \lambda_2$) is less than the top-left entry of J_1, it follows that the expression in the large parentheses is less than $-\lambda_2$, which is negative. This implies that near $(w_1, h(w_1))$, the stable manifold has a branch in the NW, and another branch

18

in the SE region. Backward integration of solutions starting near the saddle and on W_s in these regions, shows that they must either exit these regions along the boundary of Ω_{red} (because backward-time solutions cannot exit via the trapping regions), or they must converge to a steady state. However, there are no steady states to the NW of the saddle in the NW region, nor to the SE of the saddle in the SE region. Thus, the stable manifold W_s must intersect the boundary of Ω_{red} in two points, one on the line $x_1 = qw$, and the other on the line $w = s^0$. Next, we focus on the location of the unstable manifold W_u of the saddle, whose location is determined by r_2. We claim that:

$$h'(w_1) < r_2 < q.$$

To see this we consider the first equation in the eigenvalue equation for λ_2, by solving for r_2, once again recalling the expression of the derivative of the function h:

$$r_2 = \frac{\lambda_2}{Qg(s^0 - w_1)} + h'(w_1),$$

from which the first inequality follows because $\lambda_2 > 0$. The second equation in the eigenvalue equation yields that:

$$r_2 = q\frac{h(w_1)f'(p^*)}{h(w_1)f'(p^*) + \lambda_2},$$

and again, since $\lambda_2 > 0$, we find that the second inequality holds, as claimed.

We can now fully assemble the phase portrait presented in Figure 4. The stable manifold W_s of the saddle has a branch in the first and second trapping regions. Solutions starting on these branches must converge to $(0,0)$, and $(w_2, h(w_2))$ respectively. In fact, it is not hard to see that all solutions in the first trapping region converge to $(0,0)$, whereas solutions in the second and third trapping region converge to $(w_2, h(w_2))$. The fate of solutions starting in the NW and SE regions depends on their initial location relative to the separatrix W_s: They converge to $(0,0)$ if they start below W_s, but to $(w_2, h(w_2))$ if they start above W_s, and this occurs because they must enter one of the trapping regions first. $\square$

The asymptotic behavior of the solutions of system $(38) - (39)$ described in Lemma 1 can be translated into the asymptotic behavior of the solutions of system $(35) - (36)$, and combining this with the theory of asymptotically autonomous systems, see Appendix F in [25], the asymptotic behavior of the transformed system $(31) - (34)$ can be obtained as well. In turn, this determines the behavior of the scaled system $(25) - (28)$, from which Theorem 2 follows immediately.

More general digestion networks

To see why the tragedy continues to hold for more general digestion network, we consider solutions of system $(6) - (11)$, for which it is easily verified that the variable:

$$m = s + p + e + 2c + x_1 + x_2,$$

still satisfies equation (21), implying that the family of compact sets Ω_ϵ, defined earlier, is forward invariant for system $(6) - (11)$, for all $\epsilon \geq 0$, when **H2** holds. Consequently, the proof of Theorem 3 remains valid for the above chemostat model $(6) - (11)$. Indeed, the first proof only crucially depends on the dynamics of x_1 and x_2 to show that $x_1(t)$ converges to zero, after which the convergence of e, p and x_2 is obtained by elementary comparison arguments. For the digestion network presented here, the dynamics of x_1 and x_2 remain unchanged, hence we can still conclude that $x_1(t)$ converges to zero. After that, it follows from a comparison argument that $e + c$ converges to zero, and then similarly that p and x_2 converge to zero as well. One could also easily adapt the steps of the second proof to obtain the same conclusion.

References

[1] K.L. Asfahl, and M. Schuster, Social interactions in bacterial cell-cell signaling, FEMS Microbiol. Rev. 41(1), p.92-107, 2017.

[2] K.L. Asfahl, J. Walsh, K. Gilbert, and M. Schuster, Non-social adaptation defers a tragedy of the commons in Pseudomonas aeruginosa quorum sensing, ISME J. 9, p. 1734-1746, 2015.

[3] J.S. Chuang, O. Rivoire, and S. Leibler, Simpson's paradox in a synthetic microbial system, Science 323(5911), p.272-275, 2009.

[4] R. Cressman, and Y. Tao, The replicator equation and other game dynamics, PNAS 111, p. 10810-10817, 2014.

[5] A.A. Dandekar, S. Chugani, and E.P. Greenberg, Bacterial quorum sensing and metabolic incentives to cooperate, Science 338(6104), p.264-266, 2012.

[6] S.P. Diggle, A.S. Griffin, G.S. Campbell, and S.A. West, Cooperation and conflict in quorum-sensing bacterial populations, Nature 450(7168), p.411-414, 2007.

[7] F. Fiegna, and G.J. Velicer, Competitive fates of bacterial social parasites: persistence and self-induced extinction of *Myxococcus xanthus* cheaters, Proc. Biol. Sci. 270(1523), p.1527-1534, 2003.

[8] F. Fiegna, Y.T. Yu, S.V. Kadam, and G.J. Velicer, Evolution of an obligate social cheater to a superior cooperator, Nature 441(7091), p.310-314, 2006.

[9] K.R. Foster, K. Parkinson, and C.R. Thompson, What can microbial genetics teach sociobiology?, Trends Genet. 23(2), p. 74-80, 2007.

[10] K.R. Foster, S. Shaulsky, J.E. Strassmann, D.C. Queller, and C.R. Thompson, Pleiotropy as a mechanism to stabilize cooperation, Nature 431(7009), p.693-696, 2004.

[11] S.A. Frank, Foundations of social evolution, Princeton University Press, Princeton, NY, 1998.

[12] J. Gore, H. Youk, and A. van Oudenaarden, Snowdrift game dynamics and facultative cheating in yeast, Nature 459(7244), p.253-256 (2009).

[13] D. Greig, and M. Travisano, The Prisoner's Dilemma and polymorphism in yeast SUC genes, Proc. Biol. Sci. 271 Suppl. 3: S25-26, 2004.

[14] A.S. Griffin, S.A. West, and A. Buckling, Cooperation and competition in pathogenic bacteria, Nature 430 (7003), p.1024-1027, 2004.

[15] G.R. Hardin, The Tragedy of the Commons, Science 162 (3859), p. 1243-1248, 1968.

[16] G.R. Hardin, Collective action as an agreeable n-prisoner's dilemma, Behav. Sci. 16, p. 472-481, 1971.

[17] R. Kummerli, and S.P. Brown, Molecular and regulatory properties of a public good shape the evolution of cooperation, Proc Natl Acad Sci USA 107, p. 18921-18926, 2010.

[18] L. Lehmann, and L. Keller, The evolution of cooperation and altruism–a general framework and a classification of models, J. Evol. Biol. 19(5), p. 1365-1376, 2006.

[19] W. F. Lloyd, Two Lectures on the Checks to Population (Oxford Univ. Press, Oxford, Eng- land, 1833), reprinted (in part, in: Population, Evolution, and Birth Control, G. Hardin, Ed. (Freeman, San Francisco, 1964), p. 37.

[20] R.C. MacLean, and I. Gudelj, Resource competition and social conflict in experimental populations of yeast, Nature 441(7092), p. 498-501, 2006.

[21] S,J. Martin, M. Beekman, T.C. Wossler, and F.L. Ratnieks, Parasitic Cape honeybee workers, *Apis mellifera capensis*, evade policing, Nature 415(6868), p.163-165, 2002.

[22] P.B. Rainey,and K. Rainey, Evolution of cooperation and conflict in experimental bacterial populations, Nature 425 (6953), p.72-74, 2003.

[23] D.J. Rankin, K. Bargum, and H. Kokko, The tragedy of the commons in evolutionary biology, Trends Ecol. Evol. 22(12), p.643-651, 2007.

[24] K.M. Sandoz, S.M. Mitzimberg, and M. Schuster, Social cheating in *Pseudomonas aeruginosa* quorum sensing, Proc. Natl. Acad. Sci. USA 104(40), p. 15876-15881, 2007.

[25] H.L. Smith, and P. Waltman, The Theory of the Chemostat (Dynamics of Microbial Competition), Cambridge University Press, 1995.

[26] A.J. Waite, and W. Shou, Adaptation to a new environment allows cooperators to purge cheaters stochastically, Proc. Natl. Acad. Sci. USA 109(47), p.19079-19086, 2012.

[27] M. Wang, A.L. Schaefer, A.A. Dandekar, and E.P. Greenberg, Quorum sensing and policing of *Pseudomonas aeruginosa* social cheaters, Proc. Natl. Acad. Sci. USA 112(7), p. 2187-2191(2015).

[28] S.A. West, S.P. Diggle, A. Buckling, A. Gardner, and A.S. Griffins, The social lives of microbes. Annu. Rev. Ecol. Evol. Syst. 38, p. 53-77, 2007.

[29] S.A. West, A.S. Griffin, and A. Gardner, Evolutionary explanations for cooperation, Curr. Biol. 17(16), p. 661-672, 2007.

[30] S.A. West, A.S. Griffin, A. Gardner, and S.P. Diggle, Social evolution theory for microorganisms, Nat. Rev. Microbiol. 4(8), p.597-607, 2006.

[31] J.B. Xavier, W. Kim, and K.R. Foster, A molecular mechanism that stabilizes cooperative secretions in Pseudomonas aeruginosa, Mol.Microbiol. 79, p.166-179, 2011.