Norms as a Solution to the Tragedy of the Commons: A Co-evolutionary Model

Fabian Mankat*

January 30, 2024

Abstract

This paper develops a theoretical model to study when a society can rely on the cultural transmission of norms to preserve a common-pool resource (CPR). The model utilizes evolutionary game theory to endogenize the formation of personal, social, and descriptive norms, thereby unifying existing economic theories on norm evolution. By studying this model of norm evolution in a CPR game, we introduce an additional dynamic dimension, namely the resource stock. This allows us to derive insights regarding the dynamic interplay of behavior, norms, and the resource stock.

*University of Kassel, Faculty of Economics and Management

1 Introduction

Governing shared resources is among societies' most critical challenges, with its failure often resulting in severe economic and ecological consequences [Ostrom, 2009]. The underlying challenge arises from the individuals' short-term benefits of exploitation, leading to the collective depletion and eventual erosion of the resource [Hardin, 1968]. Economists widely accepted norms as one possible solution to bridge the divergence of self- and social interests in such dilemmas [Elster, 1989, Ostrom, 2000]. Generally, scholars differentiate between different types of norms [see, e.g., Bicchieri and Dimant, 2019, Dannenberg et al., in press]. A personal (or moral) norm captures an individual's perception of what is morally right, guiding behavior through inner feelings such as guilt and self-disapproval [Nyborg, 2018]. A social norm captures a society's shared understanding of what is morally right, while a descriptive norm captures what behavior members of society generally execute [Crawford and Ostrom, 1995, Thøgersen, 2006, Farrow et al., 2017]. Individuals follow social and descriptive norms to avoid social sanctions such as disapproval by others [Voss, 2001, Fehr and Fischbacher, 2004, te Velde and Louis, 2022]. Clearly, norms may introduce incentives to abstain from self-serving behavior for social benefit and, thus, may enable resource preservation.[1] However, rather than exogenous to society, norms are cultural constructs that evolve through social interactions and transmission processes. Hence, we require a comprehensive understanding of the underlying dynamic relations and inter-dependencies of personal, social, and descriptive norms, the resource stock, and behavior to understand when norms may constitute a solution to CPR dilemmas.

This paper helps to promote such an understanding. The main contribution to the existing literature is twofold. First, to the best of our knowledge, we present the first dynamic model that endogenizes personal, social, and descriptive norm formation while accounting for their joint influences on behavior. To this end, we merge existing ideas and present an evolutionary game theoretic model of norms that can be applied to various economic problems. Second, by studying this dynamic model of norms in a CPR game, we consider an additional dynamic dimension and its interactions with norms and behavior, namely the resource pool. Studying these inter-dependencies allows us to assess our research question theoretically. By doing so, we present the first model employing the *indirect evolutionary approach* by Güth and Yaari [1992] to study the evolution of pro-social traits in a CPR game.

The model of this paper consists of a large population of individuals who recurrently interact in a binary CPR game. In line with the indirect evolutionary approach, behavior results from utility maximization. Resource dynamics depend on the share of individuals refraining from resource exploitation and the resource's natural growth. Norms evolve through horizontal and oblique cultural transmission. Horizontal cultural transmission occurs through peer-to-peer interactions, whereas oblique transmission occurs through socialization institutions. These cultural transmission processes are biased in that socially successful individuals have a more significant impact on the opinion-formation process of others.

[1] For some theoretical contributions on how norms may affect economic outcomes see Bernheim [1994], Rabin [1995], Nyborg [2000], Brekke et al. [2003], Nyborg and Rege [2003b], Akerlof and Kranton [2005], Bénabou and Tirole [2006], Andreoni and Bernheim [2009], Traxler [2010], Bénabou and Tirole [2011], Figuieres et al. [2013], d'Adda et al. [2020] among others.

Social success derives from two factors, namely material payoff and social sanctions. In addition, institutional pressure positively influences norm evolution through oblique cultural transmission.

Our analysis establishes conditions for the existence and stability of (1) a boundary equilibrium point, where all individuals have internalized the norm, behave accordingly, and the resource stock is at the maximum sustainable level, (2) an interior equilibrium point, where only some individuals have internalized the norm, only these individuals behave accordingly, and the resource is below the maximum sustainable capacity, and (3) a boundary limit cycle, where all individuals have internalized the norm, behavioral conformity concerns induce society to alternate between everyone and no-one exploiting the resource, and the resource stock exhibits a cyclic evolution. The existence of these equilibria is intriguing as it provides theoretical reasoning for the practical existence of heterogeneous behavior and attitudes regarding resource exploitation, heterogeneous outcomes across different settings, and non-vanishing resource stocks in situations where the dynamic system never seems to approach a rest point.

Moreover, analyzing the stability conditions of the identified equilibria indicates that institutional pressure and the weight of social sanctions on the cultural fitness driving norm evolution play a crucial role in securing the CPR. In some instances, these two forces function as complements, whereas institutional pressure cannot be complemented in others. The latter case applies, for example, when personal sanctions play a subordinate role in determining behavioral outcomes, as this generally increases institutional pressure's effectiveness. Material costs and personal and social sanctions for norm violation play an ambiguous role as they impact cultural transmission processes as well as behavioral incentives. As a result, changes may, on the one hand, increase incentives to engage in non-exploitative behavior but, on the other hand, induce society to reach a non-favorable behavioral equilibrium in terms of norm evolution. These insights highlight the complexity of the dynamic system, underscoring the importance of understanding norm-driven behavior in a dynamic context while accounting for cultural change.

Our analysis contributes to the literature on CPR games that use evolutionary game theory. A significant contribution comes from Sethi and Somanathan [1996], who study a model where each agent in a population belongs to one of three strategy types: defectors, cooperators, and enforcers. Enforcers punish defectors, with the costs of punishing and being punished depending on the distribution of strategies. Over time, agents adapt their strategy by imitating materially successful others in the population. Sethi and Somanathan [1996] identify two potentially stable Nash equilibria: (1) a population of only defectors and (2) a population consisting of cooperators and enforcers. Noailly et al. [2003, 2007] and Tilman et al. [2020] provide further rationalizations for resource preservation when material payoff alone determines the adopted behavioral strategies. In parallel, Young [1993, 1996, 2015], Binmore and Samuelson [1994], Lindbeck et al. [1999] and others study similar preferences in public goods (PG) games.

We differ from that literature by focusing on norm-based behavioral motivations. Hence, rather than explicit material sanctioning through peers, we focus on indirect enforcement mechanisms such as self- and social disapproval. Osés-Eraso and Viladrich-Grau [2007] study such behavioral mechanisms in a CPR game. Individuals either cooperate or defect and adapt their strategies over time. Defection yields material gains, whereas cooperation yields social approval. This social approval increases in the share of others who cooperate, which introduces concerns for behavioral conformity. However, material gains of defection also increase in the share of cooperators, which enables the potential existence of a stable equilibrium characterized by the coexistence of cooperators and defectors. In addition, Osés-Eraso and Viladrich-Grau [2007] shows when stable homogeneous equilibria exist. Nyborg and Rege [2003a], Rege [2004], Nyborg et al. [2006] incorporate similar concerns for behavioral conformity into PG games, which effectively turn the respective situations into coordination games with full cooperation being a potentially stable equilibrium.

These works provide insights into how descriptive norms may affect behavioral incentives in dynamic settings. We expand on these contributions by also endogenizing the formation of personal and social norms. Conceptually and methodologically, our model closely relates to the literature that deals with the cultural transmission of norms using the indirect evolutionary approach.[2]

Mengel [2008] offers a notable contribution to this literature. She examines the cultural transmission and internalization process of a pro-social norm in a partially integrated society. The norm spreads through the imitation of successful individuals and institutional pressure. Individuals are recurrently matched to interact in the prisoners' dilemma. Any individual who has internalized the norm experiences internal sanctions when defecting. Partial integration increases the chances of alike individuals (in terms of norm internalization) to interact. In equilibrium, low levels of integration or high institutional pressure are necessary for strict norms to persist. In contrast, norms of intermediate strength can prevail under high levels of integration and low institutional pressure.

Since this paper examines a CPR game in which the entire society interacts, assortative matching cannot influence the situation. Therefore, this paper emphasizes the role of social sanctions as a co-determining factor for cultural fitness rather than that of assortative matching. In this respect, we closely relate to Mankat [2023], who studies the co-evolution of personal norms, social norms, and preferences for norm compliance in binary public goods games. He shows that if the social success determining norm evolution depends on material and social payoffs, an interplay of social disapproval mechanisms can explain the persistence of cooperation prescribing personal and social norms. We expand on this model by incorporating descriptive norms into the dynamic analysis. Moreover, we study norm evolution in a CPR rather than a PG game.

The rest of the paper unfolds as follows. Section 2 presents the static framework. We discuss equilibrium behavior for exogenous norms and resource stock in section 3. Section 4 presents the evolutionary framework, which we study in section 5. The results are discussed in section 6. Finally, section 7 concludes.

[2]Other theoretical studies of norm evolution in PG and CPR games encompass theories on rational socialization (PG: Bisin and Verdier [1998, 2001], Bisin et al. [2004], Tabellini [2008]; CPR: Schumacher [2009, 2015], Bezin [2019]), group selection (PG: Boyd and Richerson [1990, 2005], Bowles and Gintis [1998], Mitteldorf and Wilson [2000], Henrich [2004]; CPR: Waring et al. [2017]), peer persuasion (PG: Panebianco [2016]), and contagious cooperation (CPR: Richter et al. [2013], Richter and Grasman [2013]), among others. Methodological, our paper also closely relates to the literature that utilizes the indirect evolutionary approach in PG games to endogenize the formation of preferences for norm-adherence [see, e.g., Alger and Weibull, 2013, 2016, Traxler and Spichtig, 2011] and pro-social preferences in general [see, e.g., Bester and Güth, 1998, Guttman, 2003, 2013, Poulsen and Poulsen, 2006, Müller and von Wangenheim, 2019].

2 Static Framework

The model considers a continuum of individuals $i \in [0,1]$ who recurrently interact in a CPR game. An individual i's behavior $a_i \in \mathcal{A} = \{0,1\}$ is closely linked to the resource stock $r \in [0,1]$. In particular, each individual can exploit the resource, $a_i = 0$, or not exploit, $a_i = 1$. If an individual i chooses not to exploit the resource, she incurs material costs $c(r)$. Exploitation becomes increasingly profitable as the resource stock rises, $c'(r) > 0$. We write the share of individuals who do not exploit the resource as ψ. Hence, ψ indicates commonly executed behavior in society, the descriptive norm.

Since we study potential resource preservation, we normalize the resource stock to $r \in [0,1]$, where $r = 1$ indicates the maximum resource level that can persist if no one exhibits exploitative behavior and $r = 0$ indicates the point of no return so that the resource erodes irrespective of society's behavior. We discuss resource dynamics and how the resource stock changes depending on behavior in more detail in section 4.

Each individual i holds a personal norm $n_i \in \{0,1\}$ that indicates the behavior i considers morally appropriate. If $n_i = 1$, we say that individual i holds the *sustainability norm* and only considers non-exploitation to be morally appropriate. Alternatively, an individual considers all possible actions morally appropriate. We then say an individual does not hold the sustainability norm and write $n_i = 0$. If an individual believes resource exploitation is morally inappropriate, she experiences inner feelings such as guilt and loss of self-esteem if she behaves contrarily. Personal sanctions from personal norm violation correspond to $(a_i - 1)pn_i$, where $p > 0$.

In line with Cooter [1998], we define the social norm by the distribution of personal norms. Therefore, let ϕ be the share of individuals i with personal norm $n_i = 1$. If ϕ is large, then many individuals agree that non-exploitation is the only morally right thing to do, and we say that the social norm for non-exploitation is strong. If an individual exploits the resource, she is subject to social sanctions in the form of disapproval, $s(\phi, \psi)$. These social sanctions depend on the social norm ϕ as well as the descriptive norm ψ. First, the stronger the social norm is, the more individuals disapprove of exploitative behavior, and, thus, the greater are social consequences for exploiting the resource; $s'_\phi(\phi, \psi) > 0$. Second, the more individuals in society adhere to the social norm, the greater are social sanctions from non-adherence; $s'_\psi(\phi, \psi) > 0$.

3 Equilibrium Behavior

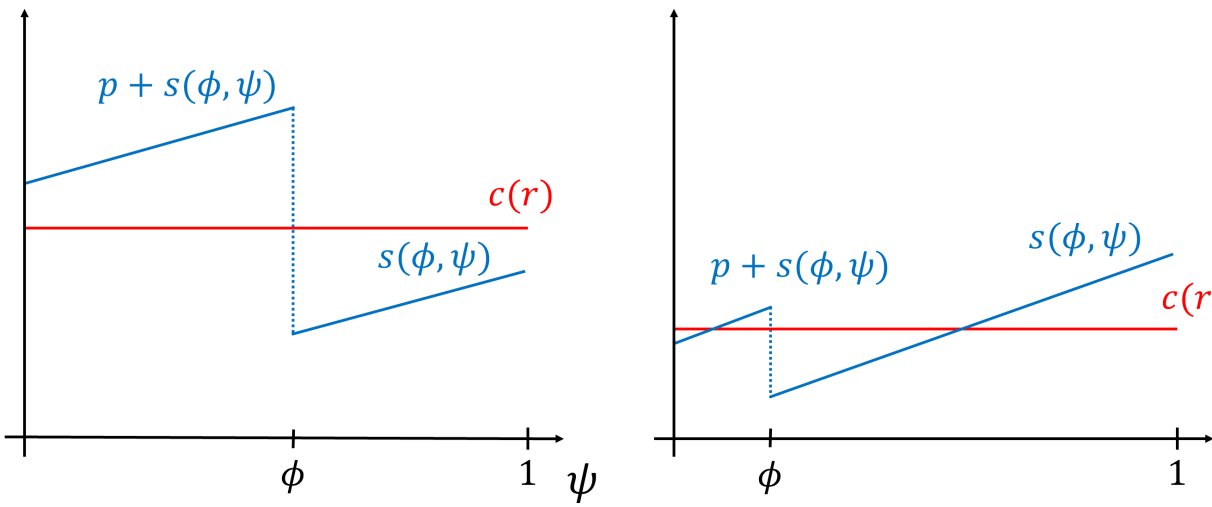
This section analyzes equilibrium behavior for any exogenous social norm ϕ and resource stock r. An individual i's behavior depends on her utility function, which is, in turn, subject to material payoff, personal sanctions, and social sanctions.

Definition 3.1 (Utility). $u_i(a_i, n_i, \phi, \psi, r) = a_i(n_i p + s(\phi, \psi) - c(r))$.

Rather than employing the Nash equilibrium concept, we require a *behavioral equilibrium* to be an evolutionary stable state. The underlying idea is that behavior itself is subject to a (very fast) evolutionary process driven by utility improvements (i.e., best-response dynamics, pairwise comparison dynamics). Let σ_n be the share of individuals i with personal norm

	Equilibrium Behavior	Existence Condition
1.	$(\sigma_1^*, \sigma_0^*) = (0, 0)$, $\psi^* = 0$	$s(\phi, 0) + p < c(r)$
2.	$(\sigma_1^*, \sigma_0^*) = (1, 0)$, $\psi^* = \phi$	$s(\phi, \psi) < c(r) < s(\phi, \psi) + p$
3.	$(\sigma_1^*, \sigma_0^*) = (1, 1)$, $\psi^* = 1$	$c(r) < s(\phi, \psi)$

Table 1: equilibrium behavior



(a) $\Psi^*(r, \phi) = \{\phi\}$ (b) $\Psi^*(r, \phi) = \{0, \phi, 1\}$

Figure 1: equilibrium behavior

$n_i = n$ who do not exploit the resource. It follows that $\psi = \phi\sigma_1 + (1-\phi)\sigma_0$. Throughout, we indicate an equilibrium share of non-exploitation by ψ^*. Moreover, the set $\Psi^*(r, \phi)$ contains all equilibrium shares ψ^* at some resource stock r and social norm ϕ. Table 1 presents all possible behavioral equilibria as well as the conditions under which they exist. There are three potential behavioral equilibria: (1) everyone exploits the resource, (2) any individual exploits the resource if and only if she does not hold the sustainability norm, and (3) no individual exploits the resource. Since the existence conditions are not mutually exclusive, different behavioral equilibria potentially co-exist. Moreover, the existence conditions cannot all be violated simultaneously, implying that $\Psi^*(r, \phi)$ is always non-empty. Below, we discuss the results using figure 1.

Figure 1 depicts two cases for which $s(\cdot)$ is linear in ψ. The red curve $c(\cdot)$ indicates the costs of non-exploitation at the respective resource stock r. The blue curve indicates personal and social sanctions at any exploitation level ψ for the next individual willing to refrain from exploiting the resource. The first ϕ individuals are subject to personal and social sanctions, whereas the later $1 - \phi$ individuals are solely subject to social sanctions. This generates the in-continuity at ϕ. At any ψ such that the blue curve lies above (below) the red curve, there must be an individual currently (not) exploiting the resource but would prefer to behave otherwise. For example, consider sub-figure 1a and any $\psi < \phi$. An individual who holds the sustainability norm exploits the resource (since $\psi < \phi$). She incurs personal and social sanctions of $s(\phi, \psi)+p$ while only avoiding material costs of $c(r)$. Consequently, she prefers to change her behavior, inducing ψ to rise. Following this line of argument, it becomes apparent why the behavioral equilibria in table 1 exist under the stated conditions. Any intersection of the blue and red curve not listed in table 1 must occur in one of the increasing segments of the blue curve. Sub-figure 1b provides an example of such intersections. Following the above reasoning, such intersections constitute Nash equilibria but cannot be evolutionary stable..

4 Evolutionary Framework

Next, we turn to the evolutionary framework that enables us to study the coevolution of behavior ψ, social norm ϕ, and resource stock r. Throughout, we write the solution of the dynamic system at time t with initial population state $\rho = (r, \phi, \psi)$ as $\xi(\rho, t)$.

Assumption 4.1. (Solution to the Dynamic System) For any population profile $\rho = (r, \phi, \psi)$, the solution $\xi(\rho, t) : [0, 1]^3 \times \mathbb{R}_{\geq 0} \to [0, 1]^3$ satisfies:

1. $\xi(\rho, 0) = (r, \phi, \psi^*)$ and
2. $\xi(\rho, t) = \xi(\lim_{\tilde{t}\to t} \xi(\rho, \tilde{t}), 0)$.

In line with the indirect approach proposed by Güth and Yaari [1992], the first condition of assumption 4.1 formalizes the idea that behavior always reaches an equilibrium before changes in the social norm and resource stock occur. The underlying intuition is that at any population state for which behavior is not in equilibrium, society moves into the nearest behavioral equilibrium at infinite speed. Section 4.1 specifies which behavioral equilibrium society moves into. Given that we can rewrite any solution $\xi(\rho, t)$ as the initial solution that starts at that state $\xi(\rho, t)$, $\xi(\rho, t) = \xi(\xi(\rho, t), 0)$, condition one implies that behavior is always in equilibrium along $\xi(\rho, t)$.[3] This introduces possible discontinuities of $\xi(\rho, t)$. If the solution approaches some state for which behavior is not in equilibrium, $\lim_{\tilde{t}\to t} \xi(\rho, \tilde{t}) = (\hat{r}, \hat{\phi}, \hat{\psi})$ s.t. $\hat{\psi} \notin \Psi^*(\hat{r}, \hat{\phi})$, then it must experience an abrupt change at t for behavior to be in equilibrium at t; $\xi(\rho, t) \neq (\hat{r}, \hat{\phi}, \hat{\psi}) = \lim_{\tilde{t}\to t} \xi(\rho, \tilde{t})$.

Condition two of assumption 4.1 implies that if the solution approaches such a discontinuity, $\lim_{\tilde{t}\to t} \xi(\rho, \tilde{t}) = (\hat{r}, \hat{\phi}, \hat{\psi})$ s.t. $\hat{\psi} \notin \Psi^*(\hat{r}, \hat{\phi})$, the dynamic system behaves as if society reaches the limit $(\hat{r}, \hat{\phi}, \hat{\psi})$ and then immediately jumps into state $\xi((\hat{r}, \hat{\phi}, \hat{\psi}), 0) = (\hat{r}, \hat{\phi}, \hat{\psi}^*)$, for which behavior is in equilibrium $\hat{\psi}^*$. Although conceptually, we can think of society being in $(\hat{r}, \hat{\phi}, \hat{\psi})$ in the blink of a moment, formally, the solution $\xi(\rho, t)$ never actually reaches this point.

[3] We can easily show that $\xi(\rho, t) = \xi(\xi(\rho, t), 0)$ indeed holds. If $\xi(\rho, t)$ is continuous at t, then $\xi(\rho, t) = \lim_{\tilde{t}\to t} \xi(\rho, \tilde{t})$. Hence, condition two of assumption 4.1 implies that $\xi(\xi(\rho, t), 0) = \xi(\lim_{\tilde{t}\to t} \xi(\rho, \tilde{t}), 0) = \xi(\rho, t)$. At any discontinuity t s.t. $\lim_{\tilde{t}\to t} \xi(\rho, \tilde{t} = (\hat{r}, \hat{\phi}, \hat{\psi})$, $\xi((\hat{r}, \hat{\phi}, \hat{\psi}), 0) = (\hat{r}, \hat{\phi}, \hat{\psi}^*)$. $\xi((\hat{r}, \hat{\phi}, \hat{\psi}), 0) = (\hat{r}, \hat{\phi}, \hat{\psi}^*) = \xi((\hat{r}, \hat{\phi}, \hat{\psi}^*), 0)$ if $\xi((r, \phi, \psi^*), 0) = (r, \phi, \psi^*)$. This last condition holds due to assumption 4.2.

4.1 Behavioral Evolution

This section discusses the evolution of behavior along the solution $\xi(\rho, t)$. Section 3 already established which behavioral equilibria exist at any resource stock r and social norm ϕ. Hence, it remains for us to identify which behavioral equilibrium society reaches.

As a starting point, consider some population profile $\rho = (r, \phi, \psi^*)$ for which behavior is in equilibrium. If society experiences marginal changes in the social norm and resource stock, the behavioral equilibrium persists, and we expect society to remain in it. For example, suppose society is at population state $\rho = (r, \phi, \psi^*)$ for which $s(\phi, \psi^*) < c(r) < s(\phi, \psi^*) + p$. Hence, the behavioral equilibrium corresponds to non-exploitation by all and only norm holders, $\psi^* = \phi$. For marginal changes in the resource and social norm to $\hat{r}$ and $\hat{\phi}$ respectively, the mixed behavioral equilibrium $\hat{\psi}^* = \hat{\phi}$ still exists, $s(\hat{\phi}, \hat{\phi}) < c(\hat{r}) < s(\hat{\phi}, \hat{\phi}) + p \Rightarrow \hat{\phi} \in \Psi^*(\hat{r}, \hat{\phi})$. Since the previous cooperation level $\psi^* = \phi$ is very close to the new behavioral equilibrium $\hat{\psi}^* = \hat{\phi}$, it is in its basin of attraction. Consequently, we expect society to coordinate into the behavioral equilibrium $\hat{\psi}^* = \hat{\phi}$. Hence, the evolution of equilibrium behavior follows that of the social norm.

Alternatively, suppose that either all individuals prefer to exploit the resource or not, $c(r) > s(\phi, \psi^*) + p$ or $c(r) < s(\phi, \psi^*)$. Hence, $\psi^* \in \{0, 1\}$. Marginal changes in the resource stock and social norm leave the behavioral equilibrium ψ^* undisturbed, and we expect equilibrium behavior to not change. Under the condition that r and ϕ are continuous with respect to time, which we will see to be the case in the following sections, we can define behavioral dynamics for any $\rho \in \{(r, \phi, \psi)\}_{\psi \in \Psi^*(r,\phi)}$ as follows.

Definition 4.1 (Behavioral Dynamics in Equilibrium)**.**

$$\dot{\psi}^* = \begin{cases} \dot{\phi} & \text{if } s(\phi, \psi^*) < c(r) < s(\phi, \psi^*) + p \\ 0 & \text{else} \end{cases}$$

Throughout, we understand all time derivatives (functions of change, dot-variables) as right-sided derivatives. Hence, they express how the variable changes as time proceeds.[4]

Definition 4.1 specifies the evolution of equilibrium behavior at any continuous point of the solution $\xi(\rho, t)$. As previously discussed, equilibrium behavior along the solution $\xi(\rho, t)$ is discontinuous if it approaches some $\hat{\rho} = (\hat{r}, \hat{\phi}, \hat{\psi})$ for which behavior is not in equilibrium, $\lim_{\tilde{t} \to t} \xi(\rho, \tilde{t}) = (\hat{r}, \hat{\phi}, \hat{\psi})$ s.t. $\hat{\psi} \notin \Psi^*(r, \phi)$. At such a discontinuity, the dynamic system transits between different behavioral equilibria. Recall that condition two of assumption 4.1 implies that at any such discontinuity, the system behaves as if it reaches the limit $\lim_{\tilde{t} \to t} \xi(\rho, \tilde{t})$ and then jumps into a population profile for which behavior is in equilibrium, $\xi(\lim_{\tilde{t} \to t} \xi(\rho, \tilde{t}), 0) = (r, \phi, \psi^*)$. Hence, to describe the transition between behavioral equilibria along the solution $\xi(\rho, t)$, it suffices to specify which behavioral equilibrium the dynamic system reaches at any $\rho = (r, \phi, \psi)$ for which behavior is (possibly) not in equilibrium. Assumption 4.2 captures the necessary specifications.

Assumption 4.2 (Transition to Behavioral Equilibrium)**.** For any population profile $\rho = (r, \phi, \psi)$ and $\psi^* \in \Psi^*(r, \phi)$, $\xi(\rho, 0) = (r, \phi, \psi^*)$ if

[4] We do so since the solution $\xi(\rho, t)$ is not perfectly differentiable, but right side differentiable.

1. $\psi^* > \psi$ and $\exists \epsilon > 0$ s.t. $s(x,\phi) + p1_{\leq\phi}(x) \geq c(r)\ \forall x \in (\psi - \epsilon, \psi^*) \cap [0,1]$ and
2. $\psi^* < \psi$ and $\exists \epsilon > 0$ s.t. $s(x,\phi) + p1_{\leq\phi}(x) \leq c(r)\ \forall x \in (\psi^*, \psi + \epsilon,) \cap [0,1]$.

The intuitive reasoning of assumption 4.2 closely follows that of the behavioral equilibria in section 3. Recall figure 1 and note that $c(r)$ and $s(x,\phi)+p1_{\leq\phi}(x)$ correspond to the red and blue curves, respectively. For illustration purposes, we focus on condition 1 of assumption 4.2. The underlying reasoning for condition 2 works analogously. Consider any starting point ψ so that for all levels of non-exploitation x below the equilibrium share ψ^*, $x < \psi^*$, and sufficiently close to ψ, $x > \psi - \epsilon$, the blue curve $s(x,\phi) + p1_{\leq\phi}(x)$ lies above the red curve $c(r)$. By similar reasoning as in section 3, the share of non-exploiting individuals must increase at any such x. Hence, behavior reaches the behavioral equilibrium ψ^*. Note that for single points $x \in (\psi - \epsilon, \psi^*)$, we allow for $c(r) = s(x,\phi) + p1_{\leq\phi}(x)$. Although behavior is at rest at such x, x is not a behavioral equilibrium. Moreover, it must hold that the non-exploitation share increases for all points in some close neighborhood around x. In the presence of constant small mutations, behavior eventually surpasses x and, hence, reaches the behavioral equilibrium ψ^*.

4.2 Resource Evolution

Resource dynamics depend on aggregate behavior ψ and resource level r.

Definition 4.2 (Resource Dynamics)**.**

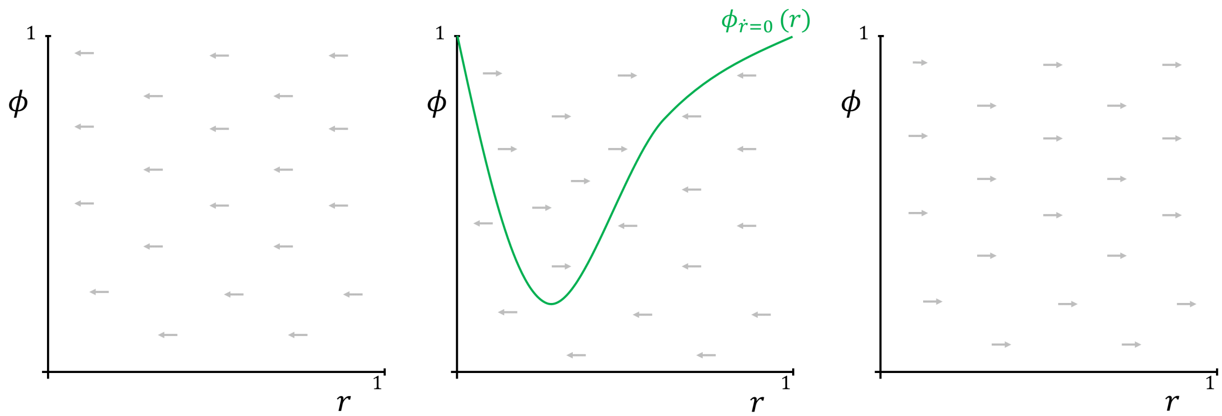
$$\dot{r} = r\delta(r) - e(1 - \psi^*),$$

where $\delta(1) = 0$, $\delta(r) > 0\ \forall r \in (0,1)$, $e(1-\psi) \in [0,\bar{e}]$, $e'(1-\psi) < 0\ \forall \psi \in [0,1]$, and $\max_{r\in(0,1]}(r\delta(r)) < \bar{e}$.

The first part of dynamics 4.2 captures the natural growth of the resource. It follows from the current resource stock r and the growth rate $\delta(r)$. If the resource is at maximum capacity, $r = 1$, then it cannot grow further, $\delta(1) = 0$. Otherwise, resource growth is positive, $\delta(r) > 0\ \forall r \in (0,1)$. For simplicity, we assume that $r\delta(r)$ has a unique maximum on the interval $(0,1]$.The second part of 4.2 captures resource extraction by society, $e(1-\psi)$. The more individuals behave sustainably, the smaller is resource extraction, $e'(x) < 0\ \forall x \in [0,1]$. We assume $\max_{r\in(0,1]}(r\delta(r)) < \bar{e}$, implying that the resource is always diminishing if everyone exploits it.[5] In the following, we discuss resource evolution at each behavioral equilibrium of section 4.1.

First, if all individuals cooperate, $\psi^* = 1$, then resource extraction is minimal, $e(1) = 0$. If the resource stock is below maximum capacity, it must be increasing, $\dot{r} > 0\ \forall r \in (0,1)$. Resource dynamics are at rest at maximum capacity, $\dot{r} = 0$ if $r = 1$. Analogously, if all individuals exploit the resource, $\psi^* = 0$, then resource extraction is at a maximum, $e(0) = \bar{e}$. in this case, the resource stock must be diminishing, $\dot{r} < 0\ \forall r \in (0,1]$. Lastly, we look at partial exploitation, $\psi^* = \phi$. Depending on the share of individuals who hold the

[5] Otherwise the situation becomes inherently uninteresting, as there is a positive resource level $\hat{r}$ below which the resource stock may never fall.



(a) $\psi^* = 0$ (b) $\psi^* = \phi$ (c) $\psi^* = 1$

Figure 2: resource evolution

sustainability norm and, thus, exploit, resource extraction $e(\phi)$ is either larger or smaller than the natural growth of the resource $r\delta(r)$. Given equilibrium behavior $\psi^* = \phi$, let the nullcline $\phi_{\dot{r}=0}(r)$ return a social norm ϕ for each resource stock r so that $r\delta(r) = e(1 - \psi^*)$. Hence, resource dynamics are at rest for any point on $\phi_{\dot{r}=0}(r)$. Suppose that at some resource stock r the social norm ϕ lies below $\phi_{\dot{r}=0}(r)$. So few individuals hold the sustainability norm and refrain from exploitation that resource extraction $e(\phi)$ outweighs natural resource growth $r\delta(r)$. The resource stock diminishes. Analogously, the resource stock rises if the social norm exceeds $\phi_{\dot{r}=0}(r)$. Figure 2 presents graphical illustrations of the above.

4.3 Norm Evolution

Following the existing literature, we focus on norm change through horizontal and oblique cultural transmission [Mengel, 2008, Mankat, 2023]. Interactions between peers and cultural imitation characterize horizontal transmission, while oblique transmission involves a learning process facilitated by socialization institutions such as schools and media. We assume that these transmission processes are biased regarding the cultural fitness of individuals. The underlying idea is that socially successful individuals have a more significant influence on the opinion formation of their peers as they are more likely to be imitated [Henrich and Gil-White, 2001] and have better prospects of occupying privileged cultural roles such as teachers or politicians [Bowles and Gintis, 1998]. In line with Traxler and Spichtig [2011] and Mankat [2023], cultural fitness results from both material rewards and social recognition.

Definition 4.3 (Cultural Fitness)**.** An individual i's cultural fitness is

$$f_i = a_i(\kappa s(\phi, \psi^*) - c(r)),$$

where $\kappa > 0$ is the weight of social sanctions on cultural fitness.

We model fitness-biased norm evolution using the well-known replicator dynamics. Hence, the resulting change in the social norm at any population profile ρ corresponds to $\phi(1-\phi)[(\sigma_1 - \sigma_0)(\kappa s(\phi, \psi^*) - c(r))$.

In addition, institutions can promote the spread of the sustainability norm by structuring interactions, stigmatizing and promoting behavior through legal norms and policies, and influencing perceptions through (public) communication and socialization institutions such as schools, universities, and media [Gintis, 2003, Mengel, 2008]. Following Gintis [2003], Mengel [2008], we capture the degree to which institutions foster norm evolution by institutional pressure $\gamma \geq 0$. Institutional pressure induces a share of individuals to adopt the sustainability norm as their personal one. Moreover, the effectiveness of institutional pressure γ is proportional to the share of individuals in society who consider the behavior morally right, the social norm.[6] Formally, the change in the social norm from institutional pressure corresponds to $\phi(1-\phi)\gamma$. Combining the above yields norm dynamics 4.4.

Definition 4.4 (Norm Dynamics)**.**

$$\dot{\phi} = v\phi(1-\phi)[(\sigma_1 - \sigma_0)(\kappa s(\phi, \psi^*) - c(r)) + \gamma],$$

where v is the relative speed of norm evolution.

As with resource dynamics 4.2, we discuss norm dynamics for each possible behavioral equilibrium ψ^*. We start with the case of either all or no individual exploiting the resource, $\psi^* \in \{0, 1\}$. Since all individuals behave equally, all individuals incur the same material costs $c(r)$ and social sanctions $s(\phi, \psi^*)$. Hence, all individuals have the same cultural fitness, implying that norm dynamics are solely driven by institutional pressure γ, $\dot{\phi} = v\phi(1-\phi)\gamma$.

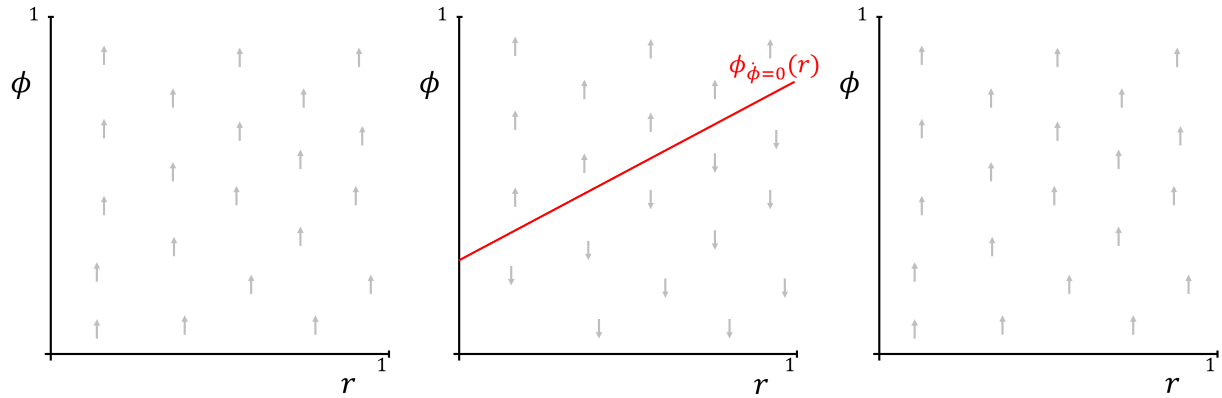
Next, the behavioral equilibrium where an individual exploits the resource if and only if she does not hold the sustainability norm, $\psi^* = \phi$. From norm dynamics 4.4, we can easily derive that norm evolution is at rest if either (1) all individuals hold the same personal norm, $\phi \in \{0, 1\}$ or (2) the difference in cultural fitness of both behavioral routines offsets institutional pressure, $\kappa s(\phi, \psi^*) + \gamma = c(r) =$. Let $\phi_{\dot{\phi}=0}(r)$ be the nullcline, which for each resource stock $r \in (0, 1]$ returns a social norm ϕ solving $\kappa s(\phi, \psi^*) + \gamma = c(r)$. Hence, norm evolution is at rest at any point on the nullcline. Note that $\phi_{\dot{\phi}=0}(r)$ is strictly increasing. A larger resource stock r corresponds to greater costs of non-exploitation $c(r)$. This requires more social sanctions $\kappa s(\phi, \phi)$ and, thus, a stronger social norm ϕ for norm evolution to be at rest. Along these lines of argument, suppose that at some resource stock r, the social norm ϕ exceeds $\phi_{\dot{\phi}=0}(r)$. Social sanctions $\kappa s(\phi, \phi)$ and institutional pressure γ outweigh the costs of contribution $c(r)$, and the social norm strengthens $\dot{\phi} > 0$. The analog holds if ϕ is below $\phi_{\dot{\phi}=0}(r)$. Figure 3 depicts the above graphically. The red lines indicate all points (r, ϕ) for which norm evolution is at rest when society exhibits equilibrium behavior $\psi^* = \phi$.

5 Evolutionary Analysis

In this section, we analyze the evolutionary model of section 4. We identify different equilibrium sets, analyze their dynamic stability, and discuss how changes in different variables may influence the situation. Throughout, we adopt the following equilibrium notion.

Definition 5.1 (Socio-Ecological Equilibrium)**.** A socio-ecological equilibrium is a closed set $\Omega \subset [0, 1]^3$ s.t. $\forall \rho \in \Omega$,

[6] Gintis [2003] and Mengel [2008] point out that proportionality is a rather conservative assumption.



(a) $\psi^* = 0$ (b) $\psi^* = \phi$ (c) $\psi^* = 1$

Figure 3: norm evolution

- $\xi(\rho, t) \in \Omega^l_{\phi=1}$ for all $t \geq 0$ and
- $\nexists \hat{\Omega} \subsetneq \Omega$ s.t. $\hat{\Omega}$ is a socio-ecological equilibrium.

Hence, a socio-ecological equilibrium corresponds to a closed set $\Omega \subset [0,1]^3$, for which the dynamic system of section 4 starting at any $\rho \in \Omega$ always remains in Ω. We require that the set Ω is minimal in the sense that it does not contain a strict subset $\hat{\Omega}$, which is itself a socio-ecological equilibrium. Our analysis identifies two different types of equilibria, namely: (1) equilibrium (or rest) points and (2) limit cycles.

5.1 Boundary Equilibrium Point

We first analyze a socio-ecological equilibrium point for which the resource stock is at maximum capacity, all individuals hold the sustainability norm, and no individual exploits the resource.

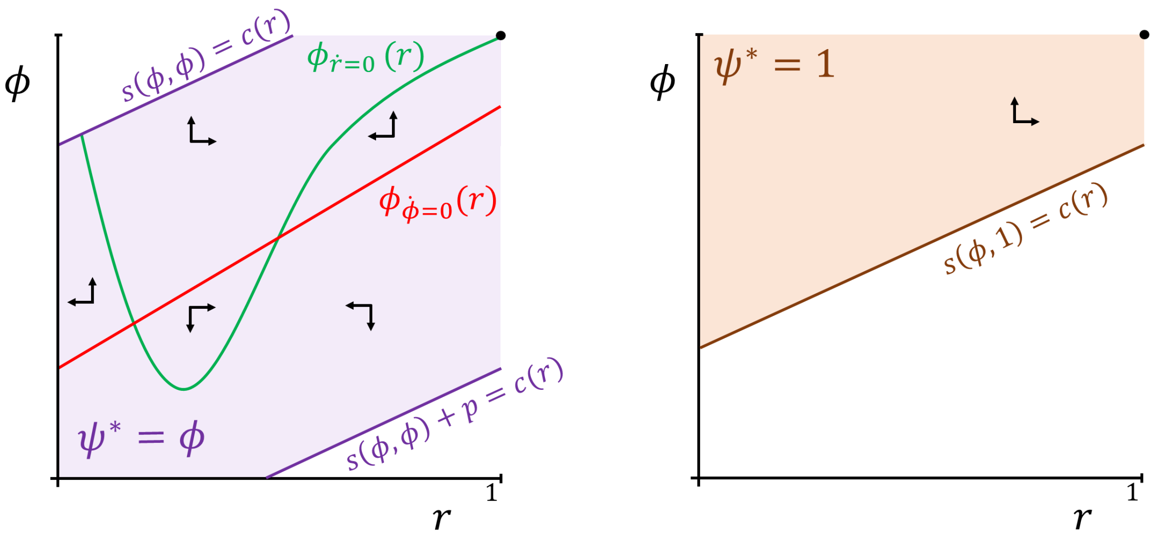
Proposition 5.1 (Boundary Socio-Ecological Equilibrium Point)**.** $(1,1,1) \in (0,1]^3$ *is a rest point if* $s(1,1) + p > c(1)$.

Proof: See appendix A.

Proposition 5.1 states that such a point is a socio-ecological equilibrium if no individual exploiting the resource is indeed a behavioral equilibrium, $s(1,1) + p > c(1) \Rightarrow \psi^* = \phi = 1$. Given the respective equilibrium behavior, $\psi^* = 1$, resource evolution is at rest if and only if the resource is at maximum capacity, $\dot{r} = 0 \Leftrightarrow r = 1$. Moreover, since all individuals hold the sustainability norm, norm evolution is at rest too, $\phi = 1 \Rightarrow \dot{\phi} = 0$.

Proposition 5.2 (Stability of Boundary Socio-Ecological Equilibrium)**.** $(1,1,1)$ *satisfying proposition 5.1 is asymptotically stable if*

1. $s(1,1) > c(1)$ *and* $\gamma > 0$ *or*
2. $s(1,1) \leq c(1) < \kappa s(1,1) + \gamma$.



(a) $s(1,1) < c(1) < \kappa s(1,1) + \gamma$ (b) $s(1,1) > c(1)$ and $\gamma > 0$

Figure 4: boundary socio-ecological equilibrium

Proof: The proposition follows from lemma 5.1 and lemma 5.2.

Proposition 5.2 discusses the stability of a boundary socio-ecological equilibrium point. Depending on how costly non-exploitation is, we can distangle it into two separate cases. We start by discussing the case of the costs of non-exploitation exceeding social sanctions, $s(1,1) \leq c(1) < s(1,1) + p$.

Lemma 5.1. *Suppose* $s(1,1) \leq c(1) < s(1,1) + p$. $(1,1,1)$ *is an asymptotically stable socio-ecological equilibrium if* $c(1) < \kappa s(1,1) + \gamma$.

Proof: See appendix A.

Lemma 5.1 states that the boundary socio-ecological equilibrium $(1,1,1)$ is asymptotically stable if social sanctions from exploitation on cultural fitness and institutional pressure outweigh material costs of non-exploitation, $\kappa s(1,1) + \gamma > c(1)$. To illustrate the underlying intuition behind this result, we focus on the case of social sanctions being strictly smaller than material costs of non-exploitation, $s(1,1) < c(1)$.[7] Hence, for all $\hat{\rho} = (\hat{r}, \hat{\phi}, \hat{\psi})$ in some neighborhood U of $(1,1,1)$, the behavioral equilibrium of partial exploitation exists, $s(\hat{\phi}, \hat{\phi}) < c(\hat{r}) < s(\hat{\phi}, \hat{\phi}) + p \Rightarrow \hat{\phi} \in \Psi^*(\hat{r}, \hat{\phi})$, and society always coordinates into it, $s(\hat{\phi}, \hat{\psi}) < c(\hat{r}) < s(\hat{\phi}, \hat{\psi}) + p \Rightarrow \xi((\hat{r}, \hat{\phi}, \hat{\psi}), 0) = (\hat{r}, \hat{\phi}, \hat{\phi})$. If U is sufficiently small, then at any population state $\hat{\rho} = (\hat{r}, \hat{\phi}, \hat{\phi}) \in U$, social sanctions from exploitation and institutional pressure outweigh material costs of non-exploitation, $\kappa s(\hat{\phi}, \hat{\phi}) + \gamma > c(\hat{r})$. The social norm strengthens, implying a return to $\phi = 1$. Since no individual exploits the resource at the perfect social norm, $\psi^* = \phi = 1$, the resource stock recovers and reaches its maximum capacity, $r = 1$. Hence, society returns to the socio-ecological equilibrium $(1,1,1)$. Figure 4a

[7] The reasoning for the alternative case of $s(1,1) = c(1)$ is similar in intuition but somewhat more elaborate.

provides a graphical illustration. Note that $\kappa s(1,1)+\gamma > c(1)$ implies that $\phi_{\dot{\phi}}(1) < 1$. Next, we turn to the case of non-exploitation being relatively cheap, $c(1) < s(1,1)$.

Lemma 5.2. *Suppose* $s(1,1) > c(1)$. $(1,1,1)$ *is a asymptotically stable if and only if* $\gamma > 0$.

Proof: See appendix A.

Lemma 5.2 states that in this case, the socio-ecological equilibrium $(1,1,1)$ is asymptotically stable if institutional pressure exists, $\gamma > 0$. Given that the costs of contribution are small, $c(1) < s(1,1)$, there is some neighborhood U of $(1,1,1)$ s.t. at each population profile $\hat{\rho} = (\hat{r},\hat{\phi},\hat{\psi})$ in that neighborhood, the behavioral equilibrium of no exploitation exists, $s(\hat{\phi},\hat{\psi}) > c(\hat{r}) \Rightarrow 1 \in \Psi^*(\hat{r},\hat{\phi})$, and society coordinates into it, $s(\hat{\phi},\hat{\psi}) > c(\hat{r}) \Rightarrow \xi((\hat{r},\hat{\phi},\hat{\psi}),0) = (\hat{r},\hat{\phi},1)$. Since all individuals refrain from resource exploitation, $\hat{\psi}^* = 1$, the resource stock increases, $\dot{\hat{r}} > 0$. If institutional pressure exists, $\gamma > 0$, the social norm also increases, $\dot{\hat{\phi}} > 0$. Hence, society evolves towards the socio-ecological equilibrium $(1,1,1)$. Figure 4b presents a graphical illustration.

Note that positive institutional pressure, $\gamma > 0$, is necessary for asymptotic stability of the socio-ecological equilibrium $(1,1,1)$. This holds since in the absence of institutional pressure, $\gamma = 0$, the social norm is at rest if no individual exploits the resource, $\psi^* = 1 \Rightarrow \dot{\phi} = 0$. Hence, it is at rest at any population profile $\hat{\rho}$ in some neighborhood U of the socio-ecological equilibrium $(1,1,1)$. At any $\hat{\rho}$, society only experiences an increase in the resource stock, $\dot{\hat{r}} > 0$. All population profiles for which the resource stock is at maximum capacity, $\hat{r} = 1$, and non-exploitation by all individuals is a behavioral equilibrium, $s(\phi,1) > c(1)$, form a connected set of rest points, $\Omega = \{(1,\phi,1)\}_{s(\phi,1)>c(1)}$ and each point in this set is Lyapunov stable but not asymptotically stable. It appears a natural next question to ask when Ω is (part of) a dynamically stable set. We can easily show that the closure of Ω is never asymptotically stable. Suppose society experiences random mutation across Ω, eventually reaching population profile $\tilde{\rho} = (1,\tilde{\phi},1)$ at the border of Ω, $s(\tilde{\phi},1) = c(\tilde{r})$. No-exploitation no longer constitutes a behavioral equilibrium. Hence, society coordinates into a new behavioral equilibrium, $\psi^* < 1$, at which the resource stock must be decreasing, $r = 1 \wedge \psi^* < 1 \Rightarrow \dot{r} < 0$. The dynamic system evolves away from Ω.

The above results indicate that increases in institutional pressure γ, the weight of social sanctions on cultural fitness κ, and personal sanctions p unambiguously favor resource preservation through a boundary socio-ecological equilibrium. Larger p favors its existence, whereas larger κ and γ favor its stability. The effect of social sanctions at the perfect social norm and no exploitation $s(1,1)$ and costs of non-exploitation at maximum resource capacity $c(1)$ are not as unambiguous. Although an increase in $s(1,1)$ and a decrease in $c(1)$ favor the existence of the respective socio-ecological equilibrium, they may alter equilibrium behavior in the socio-ecological equilibrium's neighborhood. In the absence of institutional pressure, $\gamma = 0$, this may harm stability as society moves from the case of lemma 5.1 to that of lemma 5.2. However, in the presence of institutional pressure, $\gamma > 0$, the respective changes in $s(1,1)$ and $c(1)$ favor stability.

5.2 Interior Equilibrium Point

Next, we investigate socio-ecological equilibrium points for which the resource level is (potentially) incomplete, $r < 0$.

Proposition 5.3 (Interior Socio-Ecological Equilibrium Point)**.** *Any* $(r, \phi, \phi) \in (0, 1]^3$ *is a rest point if*

1. $s(\phi, \phi) < c(r) < s(\phi, \phi) + p$,
2. $\kappa s(\phi, \phi) + \gamma = c(r)$, *and*
3. $r\delta(r) = e(1 - \phi)$.

Proof: See appendix A.

Consider any point $(r, \phi, \phi) \in (0, 1]^3$ that satisfies the conditions of proposition 5.3. If only all norm holders do not exploit the resource, $\psi = \phi$, then the material costs of non-exploitation exceed social sanctions, $s(\phi, \phi) < c(r)$, but fall short of social and personal sanctions, $c(r) < s(\phi, \phi) + p$. Hence, non-exploitation by only all norm holders indeed corresponds to a behavioral equilibrium, $\psi^* = \phi$. Given this behavioral equilibrium, section 4.2 established that resource evolution is at rest for any point on the nullcline $\phi_{\dot{r}=0}(r)$. Formally, social norm ϕ and resource level r are on this nullcline if the resource growth equals resource extraction, $r\delta(r) = e(1 - \phi)$. Similarly, norm evolution is at rest on the nullcline $\phi_{\dot{\phi}=0}(r)$, which is formally equivalent to $\kappa s(\phi, \phi) + \gamma = c(r)$. Social sanctions from norm violation on cultural fitness and institutional pressure equal the material costs of non-exploitation. If both conditions hold, then (r, ϕ) corresponds to an intersection of the two nullclines, and the dynamic system as a whole is at rest. Hence, (r, ϕ, ϕ) is a socio-ecological equilibrium. Figure 5 provides multiple graphical examples of interior socio-ecological equilibrium points.

Note that such an interior socio-ecological equilibrium can only exist at some social norm ϕ if social and personal sanctions exceed social sanctions on cultural fitness and institutional pressure, which exceed social sanctions, $s(\phi, \phi) < \kappa s(\phi, \phi) + \gamma < s(\phi, \phi) + p$. This is necessary for the behavioral equilibrium of partial exploitation to exist at ϕ and norm evolution to be at rest. Graphically, the condition ensures that the nullcline $\phi_{\dot{\phi}=0}(r)$ lies in-between the behavioral boundaries $s(\phi, \phi) + p$ and $s(\phi, \phi) + p$ in figure 5. For any social norm ϕ, this condition is more likely to hold if personal sanctions p are large, as this increases the domain of points (r, ϕ) in figure 5 for which partial exploitation is a behavioral equilibrium. Moreover, the condition is more likely satisfied if the weight of social sanctions on cultural fitness κ and institutional pressure γ jointly aggregate to an intermediate value. If κ and γ are either very large ($\gamma > p \wedge \kappa > 1$ or $\gamma > p + s(1, 1)$) or very small ($\gamma = 0$ and $\kappa < 1$), the condition may not hold for any social norm. Suppose the condition does hold at some social norm ϕ. In that case, we can easily show that if the social norm lies above some lower bound ($\phi \geq \min(\phi_{\dot{r}=0}(r))$), then there are cost curves for which an interior socio-ecological equilibrium with social norm ϕ exists.

Proposition 5.4 (Stable Interior Socio-Ecological Equilibrium Point)**.** *Any* $(r, \phi, \phi) \in (0, 1]^3$ *of proposition 5.3 is asymptotically stable if*

1. $0 < \frac{-(r\delta'(r)+\delta(r))}{e'(1-\phi)} < \frac{c'(r)}{\kappa(s'_\phi(\phi,\phi)+s'_\psi(\phi,\phi))}$ *and*

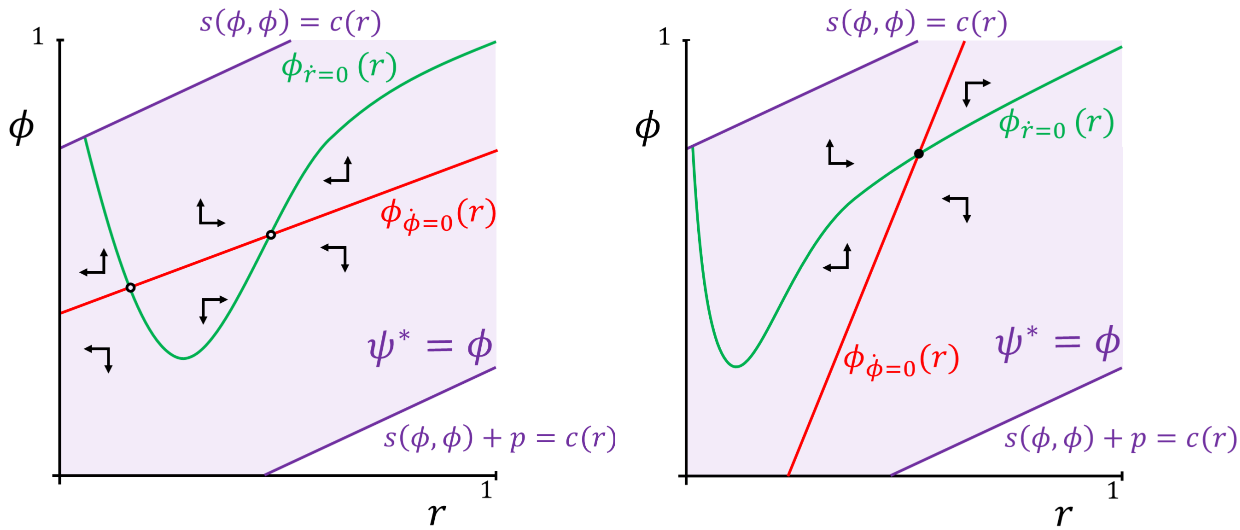
2. $v\phi(1-\phi)\kappa\left(s'_\phi(\phi,\phi)+s'_\psi(\phi,\phi)\right) < -(r\delta'(r)+\delta(r))$.

Proof: See appendix A.

Proposition 5.4 introduces conditions that ensure asymptotic stability of an interior socio-ecological equilibrium point of proposition 5.3. Recall that at any such point (r,ϕ,ϕ) individuals exploit the resource if and only if they do not hold the sustainability norm, $s(\phi,\phi) < c(r) < s(\phi,\phi)+p$. For all $\hat{\rho} = (\hat{r},\hat{\phi},\hat{\psi})$ in some neighborhood U of an interior socio-ecological equilibrium (r,ϕ,ϕ), the behavioral equilibrium of personal norm compliance exists, $s(\hat{\phi},\hat{\phi}) < c(\hat{r}) < s(\hat{\phi},\hat{\phi})+p \Rightarrow \hat{\phi} \in \Psi^*(\hat{r},\hat{\phi})$, and society coordinates into it, $s(\hat{\phi},\hat{\psi}) < c(\hat{r}) < s(\hat{\phi},\hat{\psi})+p \Rightarrow \xi((\hat{r},\hat{\phi},\hat{\psi}),0) = (\hat{r},\hat{\phi},\hat{\phi})$. Below, we focus on the dynamic system in that neighborhood U. Hence, all individuals exploit the resource if and only if they do not hold the sustainability norm throughout, $\hat{\psi}^* = \hat{\phi}$. Moreover, recall that the interior socio-ecological equilibrium corresponds to an intersection of the two nullclines $\phi_{\dot{\phi}=0}(r)$ and $\phi_{\dot{r}=0}(r)$. We can easily derive the derivatives of the two nullclines: $\phi'_{\dot{r}=0}(r) = -\frac{r\delta'(r)+\delta(r)}{e'(1-\phi)}$ and $\phi'_{\dot{\phi}=0}(r) = \frac{c'(r)}{\kappa s'(\phi)}$.

The first condition of a proposition 5.4 requires that the socio-ecological equilibrium constitutes an intersection of the two nullclines in the increasing segment of $\phi_{\dot{\phi}=0}(r)$, $0 < \frac{-(r\delta'(r)+\delta(r))}{e'(1-\phi)}$. Recall from section 4.3 that the social norm evolves away from the nullcline $\phi_{\dot{\phi}=0}(r)$. Moreover, section 4.2 shows that the resource evolves away from the decreasing segment of the nullcline $\phi_{\dot{r}=0}(r)$. Formally, the nullcline $\phi_{\dot{r}=0}(r)$ is decreasing at resource stock r if and only if an increase (decrease) in the resource stock leads to greater (lower) resource growth, $\phi'_{\dot{r}=0}(r) < 0 \Leftrightarrow \frac{dr\delta(r)}{dr} = r\delta'(r)+\delta(r) > 0$. If the socio-ecological equilibrium point was not an intersection of the two nullclines in the increasing segment of $\phi_{\dot{r}=0}(r)$, resource, as well as cultural evolution, would favor a move away from it at any $\hat{\rho}$ in some neighborhood U, implying that (r,ϕ,ϕ) is asymptotically unstable. The left intersection in figure 5a presents a respective example.

Furthermore, the first condition also requires that $\phi_{\dot{\phi}=0}(r)$ intersects $\phi_{\dot{r}=0}(r)$ from below, $-\frac{r\delta'(r)+\delta(r)}{e'(1-\phi)} < \frac{c'(r)}{\kappa\left(s'_\phi(\phi,\phi)+s'_\psi(\phi,\phi)\right)}$. For illustration purposes, suppose that this was not the case. The right intersection in figure 5a constitutes a graphical example. Recall $\frac{dr\delta(r)}{dr} = r\delta'(r)+\delta(r)$. Large values of $-(r\delta'(r)+\delta(r))$ imply that an increase (decrease) in the resource stock drastically decreases (increases) resource growth below (above) zero. Moreover, if $e'(1-\phi)$ is small, then marginal changes in the social norm, and thus the share of individuals exploiting the resource, $1-\psi^*$, only affect resource extraction slightly. Hence, at any population profile $(\hat{r},\hat{\phi},\hat{\psi})$ close to (r,ϕ,ϕ), resource evolution drives society towards some stock relatively close to r. Graphically, $\phi_{\dot{r}=0}(r)$ is steep. In conjunction with the costs of non-exploitation being relatively insensitive to changes in the resource stock, so $c'(r)$ is small, material costs of non-exploitation remain relatively steady. However, a large value of $\kappa s'(\phi)$ corresponds to social sanctions being very responsive to changes in the social norm. Hence, at any $(\hat{r},\hat{\phi},\hat{\phi})$ close to (r,ϕ,ϕ), there is a strong cultural push away from the social norm ϕ. Combining these insights implies that (r,ϕ,ϕ) is a saddle point.



(a) unstable equilibria

(b) potentially stable equilibrium

Figure 5: interior socio-ecological equilibrium

Lastly, suppose condition 1 holds. Figure 5b illustrates this case graphically. If the first condition of proposition 5.4 holds, then the socio-ecological equilibrium is asymptotically stable if resource evolution changes more drastically for marginal changes in the resource level than cultural evolution changes for marginal changes in the social norm. Condition two of proposition 5.4 captures this. Recall that norm evolution favors a move away, whereas resource evolution favors a move towards the socio-ecological equilibrium point. If condition two of proposition 5.4 holds, then, after a slight change in the resource level and social norm, the stabilizing dynamics, resource evolution, dominates the destabilizing one, norm evolution. The dynamic system spirals inwards. Alternatively, if the second condition in the lemma is violated, the evolutionary force driving society away from the socio-ecological equilibrium dominates, and the dynamic system spirals outwards.

The above results offer insights into when society may preserve a shared resource through an interior socio-ecological equilibrium. Recall that an interior socio-ecological equilibrium more likely exists for large personal sanctions p. Moreover, the weight of social sanctions on cultural fitness κ and institutional pressure γ must balance the costs of non-exploitation $c(r)$. Proposition 5.4 sheds light onto when an existing socio-ecological equilibrium is likely asymptotically stable, namely if culture evolves relatively slow (v is small), social sanctions are relatively non-responsive to changes in the social and descriptive norms (s'_ϕ and s'_ψ are small), material costs are relatively responsive to changes in the resource stock ($c'(r)$ is large), and the weight of social sanctions on cultural fitness is small (κ is small). The last bit implies that an interior socio-ecological equilibrium is more likely asymptotically stable if cultural transmission mainly occurs through institutional pressure γ. For illustration purposes, suppose that the social sanctions do not affect cultural fitness, $\kappa = 0$. In that case, changes in the social norm do not impact norm dynamics. The social norm spreads due to institutional pressure γ and erodes due to cultural fitness differences, which entirely derives from material costs $c(r)$. These two forces balance at an interior socio-ecological equilibrium, $\gamma = c(r)$. Graphically, $\phi_{\dot{\phi}=0}(r)$ corresponds to a vertical line passing through the socio-ecological equilibrium. For resource stocks $\hat{r}$ above (below) the equilibrium value r, the material costs exceed (fall short of) institutional pressure, $c(\hat{r}) > \gamma$ ($c(\hat{r}) < \gamma$), implying a norm decrease (increase). We can show that as long as the socio-ecological equilibrium occurs in the increasing segment of $\phi_{\dot{r}=0}(r)$, it is asymptotically stable.

5.3 Boundary Limit Cycle

The last socio-ecological equilibrium we discuss corresponds to a limit cycle.[8]

Proposition 5.5 (Boundary Socio-Ecological Equilibrium Limit Cycle)**.** *If* $c(0) < s(1,0) + p < s(1,1) + p \leq c(1)$, *then there is some* $\hat{r}, \check{r} \in (0,1]$ *s.t.*

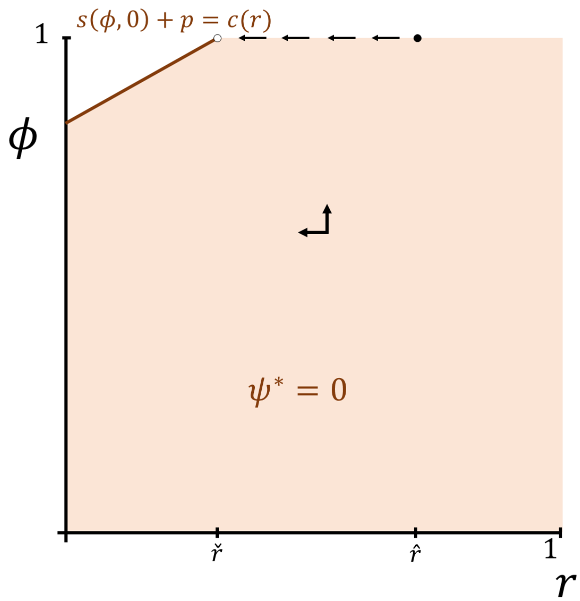
- $\forall \rho \in \Omega_{BLC} := \{(r,1,1)\}_{\check{r} \leq r < \hat{r}} \cup \{(r,1,0)\}_{\check{r} < r \leq \hat{r}}$, $\exists t > 0$ *s.t.* $\xi(\rho, 0) = \xi(\rho, t)$ *and*
- $cl(\Omega_{BLC}) = \{(r,1,1)\}_{\check{r} \leq r \leq \hat{r}} \cup \{(r,1,0)\}_{\check{r} \leq r \leq \hat{r}}$ *is a socio-ecological equilibrium.*

Proof: See appendix A.

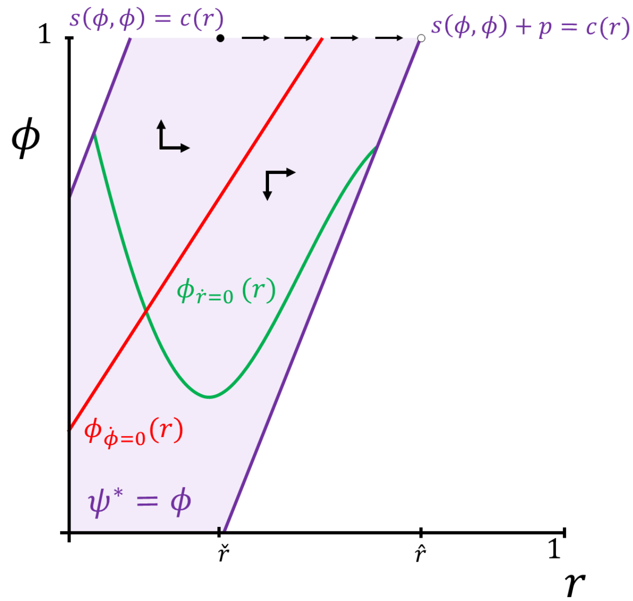
Proposition 5.5 describes a limit cycle with orbit Ω_{BLC}. Consequently, the minimal closed set containing Ω_{BLC}, $cl(\Omega_{BLC})$, constitutes a socio-ecological equilibrium. The limit cycle exists in the case of (1) the costs of non-exploitation at the full resource stock outweighing personal and social sanctions at the perfect social norm and no exploitation, $c(1) \geq p + s(1,1)$ and (2) personal and social sanctions at the perfect social norm and no cooperation outweighing the costs of non-exploitation at the minimum resource stock, $c(0) < p + s(1,0)$. Throughout the course of evolution, all individuals hold the sustainability norm as their personal norm, $\phi = 1$, implying that norm evolution is always at rest, $\dot{\phi} = 0$. Since all individuals hold the same personal norm, $\phi = 1$, society can only reach behavioral equilibria for which all individuals behave alike, $\psi^* \in \{0,1\}$. Whenever all individuals cooperate, $\psi^* = 1$, the resource stock grows, $\dot{r} > 0$. Eventually, the resource stock reaches some level $\hat{r}$ for which the costs of non-exploitation equal personal and social sanctions when no individual exploits, $c(1) \geq c(\hat{r}) = s(1,1) + p$. Non-exploitation by all individuals is no longer a behavioral equilibrium. Society transits into the behavioral equilibrium where all individuals exploit the resource, $\psi^* = 0$. The resource stock diminishes, $\dot{r} < 0$, until it reaches some level $\check{r}$ for which the costs of non-exploitation equal personal and social sanctions when all individuals exploit, $c(0) < c(\check{r}) = s(1,0) + p$. Full exploitation is no longer a behavioral equilibrium, and society transits into the behavioral equilibrium where no one exploits the resource, $\psi^* = 1$. The above process repeats.

Figure 6 presents a graphical illustration of such a limit cycle in the case of personal sanctions exceeding differences in social sanctions for all and no individuals exploiting the resource, $s(1,1) - s(0,1) < p$. Sub-figures 6a and 6b depict norm and resource evolution at each behavioral equilibrium ψ^*. Sub-figure 6c depicts the whole dynamic system in a three-dimensional graph.

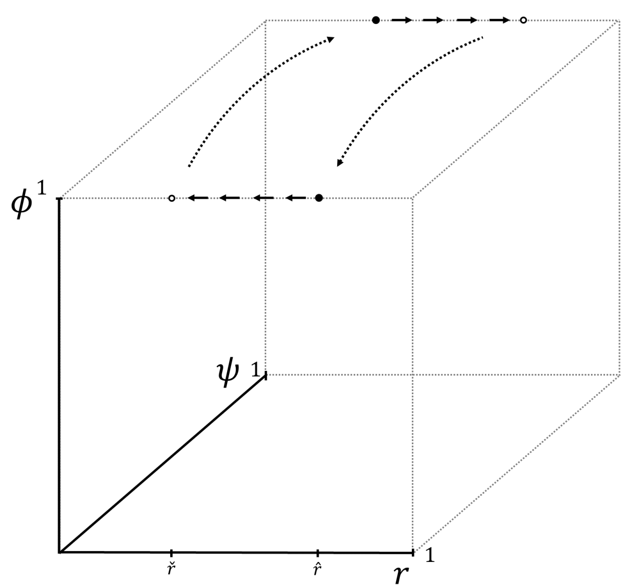
[8] Note that further limit cycles are possible, depending on the exact specifications of the dynamic system. We require extensive information about the dynamic system to ensure the existence and stability of such limit cycles. Due to the general approach taken in this paper, the detailed investigation of these limit cycles is beyond the scope of this paper.



(a) $\psi^* = 0$



(b) $\psi^* = \phi$



(c) three dimensional illustration

Figure 6: boundary limit cycle

Note that we can easily rephrase proposition 5.5 to state that the limit cycle exists only if the costs of contribution are relatively more responsive to extreme changes in the resource level than social sanctions are to extreme changes in behavior, $s(1,1) - s(1,0) < c(1) - c(0)$. Hence, the limit cycle exists only if behavioral conformity plays a relatively small role in determining social sanctions. Moreover, note that it can only exist if the boundary socio-ecological equilibrium point of proposition 5.1 does not exist.

Proposition 5.6 (Stability of a Boundary Socio-Ecological Equilibrium Limit Cycle)**.** *Consider any Ω_{BLC} of proposition 5.5.*

1. *$\forall\gamma \geq 0 \exists\kappa > 0$ s.t. $cl(\Omega_{BLC})$ is asymptotically stable,*
2. *$\forall\kappa \geq 0 \exists\gamma > 0$ s.t. $cl(\Omega_{BLC})$ is asymptotically stable,*
3. *If $(\kappa, \gamma) = (\hat{\kappa}, \hat{\gamma})$ implies $cl(\Omega_{BLC})$ is asymptotically stable, then $(\kappa, \gamma) \geq (\hat{\kappa}, \hat{\gamma})$ implies $cl(\Omega_{BLC})$ is asymptotically stable.*

Proof: See appendix A.

The above proposition provides insights regarding asymptotic stability of the boundary limit cycle of proposition 5.5. In essence, the limit cycle is (not) asymptotically stable if the weight of social sanctions on cultural fitness and institutional pressure are (not) sufficiently large. Conditions 1 and 2 state that there exist degrees of institutional pressure γ and a weight of social sanctions on cultural fitness κ for which the boundary limit cycle is asymptotically stable. Condition 3 implies that increasing these variables does not negatively impact asymptotic stability. The last two conditions state that for some institutional pressure and the weight of social sanctions on cultural fitness, the boundary limit cycle is not asymptotically stable, and a decrease in these variables is never favorable in terms of stability.

To see why the proposition holds, let us investigate Figure 6. Recall that in the specific example of figure 6, it is the case that $s(1,1) > s(1,0) + p$. Hence, at any population profile $\hat{\rho}$ in a sufficiently close neighborhood of the boundary limit cycle, society is either in the behavioral equilibrium of full or partial exploitation. The resource stock diminishes when all individuals exploit the resource at population profile $\hat{\rho}$, $\hat{\psi}^* = 0$. When full exploitation is no longer a behavioral equilibrium after a decrease in resource stock, society transits to the behavioral equilibrium of partial exploitation. Similarly, at any population profile $\hat{\rho}$ with equilibrium behavior of partial exploitation, we experience an increase in the resource stock. Hence, the cyclic tendencies of the resource stock remain present in the boundary limit cycle's neighborhood. The boundary limit cycle is asymptotically stable if the dynamic system approaches it in these cyclic movements. When all individuals exploit the resource, cultural transmission solely derives from institutional pressure and, hence, the social norm unambiguously increases. The dynamic system evolves towards the boundary limit cycle. If equilibrium behavior corresponds to partial exploitation, the social norm rises above the nullcline $\phi_{\dot{\phi}=0}(r)$ and diminishes below it. Greater (smaller) institutional pressure γ and weight of social approval on cultural fitness κ lower (increase) the nullcline and, hence, increase (lower) the area for which there is a move towards the boundary limit cycle.

Proposition 5.7 (Stability of a Boundary Socio-Ecological Equilibrium Limit Cycle 2)**.**

1. *If* $\kappa s(1,1) + \gamma > s(1,1)$*, then* $\exists \bar{p} > 0$ *s.t.* $(p < \bar{p} \wedge \Omega_{BLC}$ *of proposition 5.5 exists)* $\Rightarrow$ $cl(\Omega_{BLC})$ *is asymptotically stable.*

2. *If* $\gamma > 0$*, then for all* $\alpha > 0$*,* $\exists \bar{p} > 0$ *s.t.* $(p < \bar{p} \wedge \Omega_{BLC}$ *of proposition 5.5 exists)* $\Rightarrow$ $\{x \in [0,1]^3 : |x - \rho| < \alpha$ *for some* $\rho \in cl(\Omega_{BLC})\}$ *is asymptotically stable.*

Proof: See appendix A.

The above proposition formalizes the insight that smaller personal sanctions p might positively affect the stability of a boundary limit cycle. The first condition states that if institutional pressure γ and the weight of social sanctions κ are sufficiently large, then for sufficiently small proposal sanctions p, an existing boundary limit cycle is asymptotically stable. This holds as a decrease in personal sanctions p decreases the cooperation incentives of norm holders if such behavior is not optimal from a cultural fitness perspective. Reconsider figure 6. A decrease in p shifts the behavioral boundaries by moving the nullclines solving $s(\phi, \phi) + p = c(r)$ and $s(\phi, 0) + p = c(r)$ to the left. Hence, the limit cycle moves to the left, and society transits between behavioral equilibria at smaller resource stocks. When only all norm holders do not exploit the resource, then the social norm decreases (increases) at any point below (above) the nullcline $\phi_{\dot{\phi}=0}(r)$. This nullcline is unaffected by changes in personal sanctions p. However, moving the behavioral boundaries to the left increases the area for which society is above the nullcline at any neighborhood of the limit cycle and, hence, experiences a strengthening of the social norm and a move towards the limit cycle. Consequently, the decrease in personal sanctions favors asymptotic stability.

The second condition states that in the mere presence of institutional pressure, $\gamma > 0$, any arbitrarily close neighborhood of the limit cycle is asymptotically stable if personal sanctions p are sufficiently small. This insight suggests that minimal institutional pressure is very effective in stabilizing (an equilibrium set containing) the boundary limit cycle if personal sanctions p are small. The underlying intuition is that as we decrease personal sanctions p, the set of points for which the dynamic system is in the behavioral equilibrium where behavior differs between the norm-holder populations decreases. If the system exhibits such a behavioral equilibrium less often, it exhibits a behavioral equilibrium where institutional pressure γ alone determines cultural evolution more often. Hence, institutional pressure becomes more effective in stabilizing the limit cycle. Consider, for example, the limit case where personal sanctions are absent, $p = 0$. The limit cycle exists if $c(1) > s(1,1) > s(0,1) > c(0)$. In the neighborhood of the limit cycle, all individuals always behave alike, $\psi^* \in \{0, 1\}$, and norm dynamics are fully driven by institutional pressure, $\dot{\phi} = \phi(1-\phi)\gamma$. The mere existence of institutional pressure, $\gamma > 0$, drives society back to the perfect norm, thereby ensuring asymptotic stability of the limit cycle. Note that in this limit case of $p = 0$, the limit cycle is asymptotically stable if and only if institutional pressure exists.

Finally, we shed some light on how changes in the costs of non-exploitation $c(r)$ and social sanctions $s(\phi, \psi)$ might impact the stability of a boundary limit cycle. Generally, such changes are ambiguous, as two countervailing effects exist.

Lemma 5.3. *Consider two specifications of the dynamic model that differ only in social sanctions, and material costs $(\hat{s}(\phi,\psi),\hat{c}(r))$ and $(\check{s}(\phi,\psi),\check{c}(r))$. Suppose that (1) $\hat{\Omega}_{BLC}$ of proposition 5.5 exists at $(\hat{s}(\phi,\psi),\hat{c}(r))$, (2) $\check{\Omega}_{BLC}$ of proposition 5.5 exists at $(\check{s}(\phi,\psi),\check{c}(r))$, and (3) $\check{s}(\phi,\psi) \geq \hat{s}(\phi,\psi)$ and $\check{c}(r) \leq \hat{c}(r)$.*

1. *Suppose there exists some neighborhood N of 1 s.t. $\forall \phi \in N$, (a) $[\hat{c}^{-1}](\kappa\hat{s}(\phi,\phi)+\gamma) = [\check{c}^{-1}](\kappa\check{s}(\phi,\phi)+\gamma)$ and (b) $x < \kappa x + \gamma < x + p \; \forall x \in \{\hat{s}(\phi,\phi),\check{s}(\phi,\phi)\}$. There is U of $\hat{\Omega}_{BLC} \cup \check{\Omega}_{BLC}$ s.t.*
   (a) *$\{\rho \in U : \xi(\rho,0) = (r,\phi,\phi) \wedge \dot{\phi} > 0$ at $(\check{s}(\phi,\psi),\check{c}(r))\} \subseteq \{\rho \in U : \xi(\rho,0) = (r,\phi,\phi) \wedge \dot{\phi} > 0$ at $(\hat{s}(\phi,\psi),\hat{c}(r))\}$ and*
   (b) *$\{\rho \in U : \xi(\rho,0) = (r,\phi,\phi) \wedge \dot{\phi} < 0$ at $(\hat{s}(\phi,\psi),\hat{c}(r))\} \subseteq \{\rho \in U : \xi(\rho,0) = (r,\phi,\phi) \wedge \dot{\phi} < 0$ at $(\check{s}(\phi,\psi),\check{c}(r))\}$.*
2. *For all ρ s.t. $\xi(\rho,0) = (r,\phi,\phi)$ at $(\hat{s}(\phi,\psi),\hat{c}(r))$ and $(\check{s}(\phi,\psi),\check{c}(r))$, $\dot{\phi}$ is larger (smaller) at $(\hat{s}(\phi,\psi),\hat{c}(r))$ than at $(\check{s}(\phi,\psi),\check{c}(r))$.*

Proof: See appendix A.

The above lemma describes these two opposing effects. First, similar to an increase in personal sanctions p, a decrease in $c(r)$ and/or an increase in $s(\phi,\psi)$ shift the boundaries of the behavioral equilibria and, hence, the whole limit cycle to the right in figure 6. By the same reasoning as above, this effect generally harms stability of the boundary limit cycle. Second, the described changes in $c(r)$ and $s(\phi,\psi)$ favor norm evolution at any social norm ϕ and resource stock r if behavior is in the partial exploitation equilibrium $\psi^* = \phi$. The decrease in $c(r)$ and increase in $s(\phi,\phi)$ favor the cultural fitness of the norm holders and, thereby, norm evolution. At any point for which only all norm-holders do not exploit the resource in figure 6, the respective changes let all vectors of motion point further up and move the nullcline $\phi_{\dot{\phi}=0}(r)$ to the right. This effect generally favors stability of the limit cycle. Whether the changes in $c(r)$ and $s(\phi,\psi)$ are beneficial in terms of stability depends on which of these two effects dominates, which is, in turn, subject to the exact specifications of the dynamic system.

In summary, the results of this section show that resource preservation through a boundary limit cycle is possible if the costs of non-exploitation at the maximum resource level $c(1)$ are high and relatively responsive to changes in the resource stock as compared to social sanctions for behavioral non-conformity, $c(1) - c(0) > s(1,1) - s(1,0)$. Although personal sanctions p may aid the existence of a limit cycle (by ensuring that full exploitation is not a behavioral equilibrium as society approaches the point of no return), too large personal sanctions p may threaten stability of an existing one. This adverse effect of personal sanctions on stability becomes even more prominent by observing that institutional pressure γ is especially effective in stabilizing the limit cycle when personal sanctions are small. Changes in social sanctions $s(\phi,\psi)$ and material costs $c(r)$ have a more ambiguous effect. They shift the boundaries of behavioral equilibria as well as impact norm dynamics at any particular behavioral equilibrium. Finally, greater institutional pressure γ and weight of cultural fitness κ generally positively affect stability.

6 Discussion

Our analysis suggests that if the costs of non-exploitation at the maximum resource stock are small, $c(1) < s(1,1)$, society can preserve the CPR through a stable boundary socio-ecological equilibrium with minimal levels of institutional pressure, $\gamma > 0$. For intermediate costs of non-exploitation, $s(1,1)+p > c(1) \geq s(1,1)$, society requires jointly sufficiently large institutional pressure and weight of social sanctions on cultural fitness to do so, $\kappa s(1,1)+\gamma > c(1)$. However, contrary to the previous case, institutional pressure is no longer necessary for stability. If non-exploitation is very costly at the maximum resource stock, $s(1,1)+p \leq c(1)$, but relatively cheap as the resource approaches the point of no return and behavioral conformity plays a relatively minor role in determining sanctions from norm violation, $c(0) < s(1,0)+p$, then society may preserve the CPR through a stable boundary limit cycle if the weight of social sanctions on cultural fitness κ and institutional pressure γ are again sufficiently large. Alternatively, there exists an interior socio-ecological equilibrium at social norm ϕ and resource stock r if only all norm holders not exploiting the resource is a behavioral equilibrium, $s(\phi,\phi)+p > c(r) > s(\phi,\phi)$, and institutional pressure and the weight of social sanctions on cultural fitness balance the costs of contribution, $\gamma + \kappa s(\phi,\phi) = c(r)$. Society is more likely to preserve the resource through such a stable interior socio-ecological equilibrium if institutional pressure γ plays a more significant role than the weight κ in balancing material costs $c(r)$.

The discussion above indicates that, at times, institutional pressure γ and the weight of social sanctions on cultural fitness κ constitute complements in securing the CPR. In contrast, sometimes γ cannot be complemented by κ. In these latter cases, (relatively) large κ is either ineffective (at the boundary equilibrium if $s(1,1) > c(r)$) or harming stability (at the interior equilibrium point). Moreover, our results suggest that this complementarity weakens if personal sanctions p are relatively unimportant in the sense that either social sanctions $s(\phi,\psi)$ suffice to motivate non-exploitative behavior or personal sanctions p are small, as this generally renders minimal institutional pressure stabilizing boundary socio-ecological equilibria.

Note, however, that there are instances for which too much institutional pressure γ is also non-favorable in terms of resource preservation, as it may hinder the existence of an interior socio-ecological equilibrium (e.g. if $\gamma > p + s(1,1)$). If material costs at the maximum resource stock are too large for a boundary socio-ecological equilibrium to exist, $p + s(1,1) < c(1)$, and the descriptive social norm plays a relatively large role in determining total sanctions from personal and social norm violation, $s(1,0)+p < c(0)$, then society could benefit from lower levels of institutional pressure that enable an interior socio-ecological equilibrium. This line of argument easily extends to the weight of social sanctions on cultural fitness κ.

Beyond institutional pressure γ and the weight of social sanctions on cultural fitness κ, we can draw conclusions regarding the further variables. We have already touched upon personal sanctions p, for which small values may favor the stability of a limit cycle. However, larger values of p increase the cooperation incentives of norm holders, thereby favoring the existence of such a limit cycle as well as that of boundary and interior socio-ecological equilibria. Hence, personal sanctions play a somewhat ambiguous role. We can accredit a similar ambiguous role to social sanctions $s(\phi, \psi)$ and material costs of contribution $c(r)$. Changes in these variables affect behavioral boundaries. In particular, larger social sanctions $s(\phi, \psi)$ and smaller material costs $c(r)$ generally favor the existence of a boundary equilibrium point or limit cycle. However, for an existing boundary equilibrium, the magnitude of these variables has an ambiguous effect on stability. For the boundary equilibrium point, for example, the described changes are potentially non-favorable in the absence of institutional pressure and favorable in its presence. These results highlight the complexity of the dynamic system and the importance of understanding norms in a dynamic context. For example, when designing policies targeting behavioral incentives, these considerations must account for potentially adverse effects on norm dynamics.

Finally, our findings suggest that a strong sensitivity of material costs to variations in the resource stock is favorable for securing the CPR. The existence of a boundary limit cycle is contingent upon the material costs being responsive to extreme changes in the resource stock, and an existing interior socio-ecological equilibrium point is stable when marginal changes in the resource stock induce a significant alteration in material costs.

7 Concluding Remarks

This paper aims to deepen the understanding of the dynamic interactions between behavior, norms, and resources to shed light on when a society can rely on norm transmission to maintain a CPR. To this end, we identify asymptotically stable equilibrium sets of the dynamic system characterized by positive resource levels. We find that, under certain conditions, there exist (a) asymptotically stable equilibrium points where the social and descriptive norms coincide, and personal norms and behavior are either heterogeneous or homogeneous across individuals, and (b) an asymptotically stable limit cycle that exhibits non-varying personal and social norms, but herding leads to alternating descriptive norms and resource stocks. Our findings point to ambiguous roles of most variables for the existence and stability of these equilibrium sets, highlighting the complex nature of the underlying dynamic system. This emphasizes the necessity for context-specific analyses as well as acquiring a more comprehensive understanding of the dynamics at play.

Future research should complement this paper by expanding on the equilibrium analyses and discussing how societies can transition from one equilibrium point to a socially preferred one. Similarly, further research must analyze theoretically how policy measures perform when accounting for the multi-layered dynamics. Our findings suggest that the effectiveness of policies that alter different variables is ambiguous, with their implementation potentially resulting in unfavorable consequences. Lastly, the analysis should be expanded to account for other mechanisms influencing norm dynamics (e.g., cognitive dissonance, normative conformity concerns, network effects) to draw a more complete picture of the underlying dynamic system.

References

George A Akerlof and Rachel E Kranton. Identity and the economics of organizations. *Journal of Economic Perspectives*, 19(1):9–32, 2005.

Ingela Alger and Jörgen W Weibull. Homo moralis—preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302, 2013.

Ingela Alger and Jörgen W Weibull. Evolution and kantian morality. *Games and Economic Behavior*, 98:56–67, 2016.

James Andreoni and B Douglas Bernheim. Social image and the 50–50 norm: A theoretical and experimental analysis of audience effects. *Econometrica*, 77(5):1607–1636, 2009.

Roland Bénabou and Jean Tirole. Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678, 2006.

Roland Bénabou and Jean Tirole. Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855, 2011.

B Douglas Bernheim. A theory of conformity. *Journal of Political Economy*, 102(5):841–877, 1994.

Helmut Bester and Werner Güth. Is altruism evolutionarily stable? *Journal of Economic Behavior & Organization*, 34(2):193–209, 1998.

Emeline Bezin. The economics of green consumption, cultural transmission and sustainable technological change. *Journal of Economic Theory*, 181:497–546, 2019. ISSN 0022-0531.

Cristina Bicchieri and Eugen Dimant. Nudging with care: the risks and benefits of social information. *Public Choice*, pages 1–22, 2019.

Ken Binmore and Larry Samuelson. An economist's perspective on the evolution of norms. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pages 45–63, 1994.

Alberto Bisin and Thierry Verdier. On the cultural transmission of preferences for social status. *Journal of Public Economics*, 70(1):75–97, 1998.

Alberto Bisin and Thierry Verdier. The economics of cultural transmission and the dynamics of preferences. *Journal of Economic Theory*, 97(2):298–319, 2001.

Alberto Bisin, Giorgio Topa, and Thierry Verdier. Cooperation as a transmitted cultural trait. *Rationality and Society*, 16(4):477–507, 2004.

Samuel Bowles and Herbert Gintis. The moral economy of communities: Structured populations and the evolution of pro-social norms. *Evolution and Human Behavior*, 19(1):3–25, 1998.

Robert Boyd and Peter J. Richerson. Group selection among alternative evolutionarily stable strategies. *Journal of Theoretical Biology*, 145(3):331–342, 1990. ISSN 0022-5193.

Robert Boyd and Peter J Richerson. *The origin and evolution of cultures*. Oxford University Press, 2005.

Kjell Arne Brekke, Snorre Kverndokk, and Karine Nyborg. An economic model of moral motivation. *Journal of Public Economics*, 87(9-10):1967–1983, 2003.

Robert Cooter. Expressive law and economics. *The Journal of Legal Studies*, 27(S2):585–607, 1998.

Sue ES Crawford and Elinor Ostrom. A grammar of institutions. *American Political Science Review*, pages 582–600, 1995.

Giovanna d'Adda, Martin Dufwenberg, Francesco Passarelli, and Guido Tabellini. Social norms with private values: Theory and experiments. *Games and Economic Behavior*, 124: 288–304, 2020.

Astrid Dannenberg, Gunnar Gutsche, Marlene Batzke, Sven Christens, Daniel Engler, Fabian Mankat, Sophia Moeller, Eva Weingaertner, Andreas Ernst, Marcel Lumkowsky, et al. The effects of norms on environmental behavior. *Review of Environmental Economics and Policy*, in press.

Jon Elster. Social norms and economic theory. *Journal of Economic Perspectives*, 3(4): 99–117, 1989.

Katherine Farrow, Gilles Grolleau, and Lisette Ibanez. Social norms and pro-environmental behavior: a review of the evidence. *Ecological Economics*, 140:1–13, 2017.

Ernst Fehr and Urs Fischbacher. Social norms and human cooperation. *Trends in Cognitive Sciences*, 8(4):185 – 190, 2004. ISSN 1364-6613.

Charles Figuieres, David Masclet, and Marc Willinger. Weak moral motivation leads to the decline of voluntary contributions. *Journal of Public Economic Theory*, 15(5):745–772, 2013.

Herbert Gintis. Solving the puzzle of prosociality. *Rationality and Society*, 15(2):155–187, 2003.

Werner Güth and Menahem Yaari. An evolutionary approach to explain reciprocal behavior in a simple strategic game. *U. Witt. Explaining Process and Change–Approaches to Evolutionary Economics. Ann Arbor*, pages 23–34, 1992.

Joel M Guttman. Repeated interaction and the evolution of preferences for reciprocity. *The Economic Journal*, 113(489):631–656, 2003.

Joel M Guttman. On the evolution of conditional cooperation. *European Journal of Political Economy*, 30:15–34, 2013.

Garrett Hardin. The tragedy of the commons: the population problem has no technical solution; it requires a fundamental extension in morality. *Science*, 162(3859):1243–1248, 1968.

Joseph Henrich. Cultural group selection, coevolutionary processes and large-scale cooperation. *Journal of Economic Behavior & Organization*, 53(1):3–35, 2004.

Joseph Henrich and Francisco J Gil-White. The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and Human Behavior*, 22(3):165–196, 2001.

Assar Lindbeck, Sten Nyberg, and Jörgen W Weibull. Social norms and economic incentives in the welfare state. *The Quarterly Journal of Economics*, 114(1):1–35, 1999.

Fabian Mankat. Cooperation, norms, and gene-culture coevolution. *Available at SSRN 4455734*, 2023.

Friederike Mengel. Matching structure and the cultural transmission of social norms. *Journal of Economic Behavior & Organization*, 67(3-4):608–623, 2008.

Joshua Mitteldorf and David Sloan Wilson. Population viscosity and the evolution of altruism. *Journal of Theoretical Biology*, 204(4):481–496, 2000.

Stephan Müller and Georg von Wangenheim. Coevolution of cooperation, preferences, and cooperative signals in social dilemmas. *Center for European, Governance, and Economic Development Research, Discussion Paper*, (221), 2019.

Joëlle Noailly, Cees A Withagen, and Jeroen CJM Van den Bergh. Spatial evolution of social norms in a common-pool resource game. *Environmental and Resource Economics*, 36:113–141, 2007.

Joïlle Noailly, Jeroen CJM van den Bergh, and Cees A Withagen. Evolution of harvesting strategies: replicator and resource dynamics. *Journal of Evolutionary Economics*, 13: 183–200, 2003.

Karine Nyborg. Homo economicus and homo politicus: interpretation and aggregation of environmental values. *Journal of Economic Behavior & Organization*, 42(3):305–322, 2000.

Karine Nyborg. Social norms and the environment. *Annual Review of Resource Economics*, 10:405–423, 2018.

Karine Nyborg and Mari Rege. On social norms: the evolution of considerate smoking behavior. *Journal of Economic Behavior & Organization*, 52(3):323–340, 2003a.

Karine Nyborg and Mari Rege. Does public policy crowd out private contributions to public goods. *Public Choice*, 115(3):397–418, 2003b.

Karine Nyborg, Richard B Howarth, and Kjell Arne Brekke. Green consumers and public policy: On socially contingent moral motivation. *Resource and Energy Economics*, 28(4): 351–366, 2006.

Nuria Osés-Eraso and Montserrat Viladrich-Grau. On the sustainability of common property resources. *Journal of Environmental Economics and Management*, 53(3):393–410, 2007.

Elinor Ostrom. Collective action and the evolution of social norms. *Journal of Economic Perspectives*, 14(3):137–158, September 2000.

Elinor Ostrom. A general framework for analyzing sustainability of social-ecological systems. *Science*, 325(5939):419–422, 2009.

Fabrizio Panebianco. The role of persuasion in cultural evolution dynamics. *International Review of Economics*, 63(3):233–258, 2016.

Anders Poulsen and Odile Poulsen. Endogenous preferences and social-dilemma institutions. *Journal of Institutional and Theoretical Economics (JITE)/Zeitschrift für die gesamte Staatswissenschaft*, pages 627–660, 2006.

Matthew Rabin. Moral preferences, moral constraints, and self-serving biases. 1995.

Mari Rege. Social norms and private provision of public goods. *Journal of Public Economic Theory*, 6(1):65–77, 2004.

Andries Richter and Johan Grasman. The transmission of sustainable harvesting norms when agents are conditionally cooperative. *Ecological Economics*, 93:202–209, 2013.

Andries Richter, Daan van Soest, and Johan Grasman. Contagious cooperation, temptation, and ecosystem collapse. *Journal of Environmental Economics and Management*, 66(1): 141–158, 2013.

Ingmar Schumacher. The dynamics of environmentalism and the environment. *Ecological Economics*, 68(11):2842–2849, 2009.

Ingmar Schumacher. The endogenous formation of an environmental culture. *European Economic Review*, 76:200–221, 2015.

Rajiv Sethi and Eswaran Somanathan. The evolution of social norms in common property resource use. *The American Economic Review*, pages 766–788, 1996.

Guido Tabellini. The scope of cooperation: Values and incentives. *The Quarterly Journal of Economics*, 123(3):905–950, 2008.

Vera L te Velde and Winnifred Louis. Conformity to descriptive norms. *Journal of Economic Behavior & Organization*, 200:204–222, 2022.

John Thøgersen. Norms for environmentally responsible behaviour: An extended taxonomy. *Journal of Environmental Psychology*, 26(4):247–261, 2006.

Andrew R Tilman, Joshua B Plotkin, and Erol Akçay. Evolutionary games with environmental feedbacks. *Nature Communications*, 11(1):915, 2020.

Christian Traxler. Social norms and conditional cooperative taxpayers. *European Journal of Political Economy*, 26(1):89–103, 2010.

Christian Traxler and Mathias Spichtig. Social norms and the indirect evolution of conditional cooperation. *Journal of Economics*, 102(3):237–262, 2011.

Thomas Voss. *Game-theoretical perspectives on the emergence of social norms.* na, 2001.

Timothy M Waring, Sandra H Goff, and Paul E Smaldino. The coevolution of economic institutions and sustainable consumption via cultural group selection. *Ecological Economics*, 131:524–532, 2017.

H Peyton Young. The evolution of conventions. *Econometrica: Journal of the Econometric Society*, pages 57–84, 1993.

H Peyton Young. The economics of convention. *Journal of Economic Perspectives*, 10(2): 105–122, 1996.

H Peyton Young. The evolution of social norms. *Economics*, 7(1):359–387, 2015.

A Proofs

Proof of proposition 5.1. Follows from $\phi = 1 \Rightarrow \dot{\phi} = 0$, $s(1,1) + p > c(1) \Rightarrow 1 \in \Psi^*(1,1) \Rightarrow \dot{\psi}^* \in \{0, \dot{\phi}\} \Rightarrow \dot{\psi}^* = 0$, and $(\psi^* = 1 \wedge r = 1) \Rightarrow \dot{r} = 0$. □

Proof of lemma 5.1. Suppose that $s(1,1) < c(1) < s(1,1) + p$. For all $\hat{\rho} = (\hat{r}, \hat{\phi}, \hat{\psi})$ in some U of $(1,1,1)$, $s(\hat{\phi}, \hat{\psi}) < c(\hat{r}) < s(\hat{\phi}, \hat{\psi}) + p$. Hence, $\xi(\hat{\rho}, 0) = (\hat{r}, \hat{\phi}, \hat{\phi})\ \forall \hat{\rho} \in U$. Since $\hat{\psi}^* = \hat{\phi}\ \forall \hat{\rho} \in U$, we can restrict our attention to the reduced dynamic system consisting of (1) $\dot{\hat{r}} = \hat{r}\delta(\hat{r}) - e(1 - \hat{\phi})$ and (2) $\dot{\hat{\phi}} = v\hat{\phi}(1 - \hat{\phi})(\kappa\,(s'_\phi(\phi, \phi) + s'_\psi(\phi, \phi)) + \gamma - c(\hat{r}))$ to analyze asymptotic stability of $(1,1,1)$. Linearization of the reduced system around $(1,1,1)$ yields the Jacobian matrix $J = \left(\begin{smallmatrix} \delta'(1) & e'(0) \\ 0 & c(1) - \kappa s(1,1) - \gamma \end{smallmatrix}\right)$. $(1,1,1)$ is asymptotically stable if trace $tr(J) < 0$ and trace $det(J) > 0$. $tr(J) = \delta'(1) + c(1) - \kappa s(1,1) - \gamma < 0$ if $c(1) < \kappa s(1,1) + \gamma$. This holds since $\delta(r) * r$ has a unique maximum at some $r \in (0,1)$, implying that $\delta'(\hat{r})\hat{r} + \delta(\hat{r}) < 0\ \forall \hat{r} > r$ and, thus, $\delta'(1) * 1 + \delta(1) = \delta'(1) < 0$. $det(J) = \delta'(1) * (c(1) - \kappa s(1,1) - \gamma) > 0$. Hence, $(1,1,1)$ is asymptotically stable.

Next, suppose $s(1,1) = c(1) < s(1,1) + p$. $s(1,1) + p > c(1) \Rightarrow \exists U_1$ of $(1,1,1)$ s.t. $s(\phi, \psi) + p > c(r)\ \forall \rho \in U_1$. Hence, $\xi(\rho, 0) \in \{(r, \phi, 1), (r, \phi, \phi)\} \forall \rho = (r, \phi, \psi) \in U_1$. $s(1,1) + p > s(1,1) \Rightarrow \exists U_2$ of $(1,1,1)$ s.t. $s(\phi, \phi) + p > s(\phi, 1)\ \forall \rho = (r, \phi, \psi) \in U_2$. $\kappa s(1,1) + \gamma > c(1) \Rightarrow \exists U_3$ of $(1,1,1)$ s.t. $\kappa s(\phi, \phi) + \gamma > c(r)\ \forall \rho = (r, \phi, \phi) \in U_3$. Hence, $\xi(\rho, 0) = (r, \phi, \phi) \Rightarrow \dot{\phi} > 0$. Let $U \subset U_1 \cap U_2 \cap U_3$ be so small that for the reduced dynamic system of norms and resources given $\psi^* = \phi$, the solution starting at any ρ in U always remains in $U_1 \cap U_2 \cap U_3$. Such U exists due to our previous results on asymptotic stability of the reduced dynamic system (see the linearization above).

Consider any $\rho \in U$ s.t. $\xi(\rho, 0) = (r, \phi, \phi)$. $\xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, \bar{\phi})\ \forall \bar{t} > 0 \Rightarrow \lim_{t \to \infty} \xi(\rho, t) = (1,1,1)$, due to the results above. Hence, the system converges to $(1,1,1)$. Alternatively, suppose the contrary is true. Let $\bar{t} = \min\{t > 0 : \xi(\rho, t) \neq (r, \phi, \phi)\}$. $\xi(\rho, t) = (r, \phi, \phi) \in U \subset U_1\ \forall t < \bar{t} \Rightarrow \lim_{t \to \bar{t}} \xi(\rho, t) \in U_1 \Rightarrow \xi(\lim_{t \to \bar{t}} \xi(\rho, t), 0) = (\bar{r}, \bar{\phi}, 1) \Rightarrow \xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, 1)$.

Next, consider any $\rho \in U$ s.t. $\xi(\rho, 0) = (r, \phi, 1)$. $\psi^* = 1 \Rightarrow \dot{\phi} \geq 0 \wedge \dot{r} \geq 0$. $\xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, 1)\ \forall \bar{t} > 0 \Rightarrow \lim_{t \to \infty} \xi(\rho, t) = (1,1,1)$. To see why this holds, suppose $\xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, 1)\ \forall \bar{t} > 0$ and $\lim_{t \to \infty} \xi(\rho, t) \neq (1,1,1)$. Note that $(\psi^* = 1 \wedge r < 1 \Rightarrow \dot{r} > 0) \Rightarrow \lim_{t \to \infty} \xi(\rho, t) = (1, x, 1) \neq (1,1,1)$ for some $x < 1$. However, $s(1,1) = c(1) \Rightarrow s(x,1) < c(1) \Rightarrow \exists \bar{U}$ of $(1, x, 1)$ s.t. $1 \notin \Psi^*(r, \phi)\ \forall \rho \in \bar{U} \Rightarrow \xi(\rho, \bar{t}) \neq (\bar{r}, \bar{\phi}, 1)$ for some $\bar{t} > 0$. We have reached a contradiction, implying that: $\xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, 1)\ \forall \bar{t} > 0 \Rightarrow \lim_{t \to \infty} \xi(\rho, t) = (1,1,1)$. Alternatively, suppose $\xi(\rho, \bar{t}) \neq (\bar{r}, \bar{\phi}, 1)$ for some $\bar{t} > 0$. Let $\bar{t} = \min\{t > 0 : \xi(\rho, t) \neq (\bar{r}, \bar{\phi}, 1)\}$. $\xi(\rho, \tilde{t}) = (\tilde{r}, \tilde{\phi}, 1)\ \forall \tilde{t} < \bar{t} \Rightarrow \dot{\tilde{r}} \geq 0 \wedge \dot{\tilde{\phi}} \geq 0 \Rightarrow \xi(\rho, \tilde{t}) = (\tilde{r}, \tilde{\phi}, 1) \in U_1\ \forall \tilde{t} < \bar{t} \Rightarrow \lim_{\tilde{t} \to \bar{t}} \xi(\rho, \tilde{t}) \in U_1 \Rightarrow \xi(\lim_{\tilde{t} \to \bar{t}} \xi(\rho, \tilde{t}), 0) = (\bar{r}, \bar{\phi}, \bar{\phi}) \Rightarrow \xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, \bar{\phi})$.

The above implies that as soon as the dynamic system remains one behavioral equilibrium, $\xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, \bar{\phi})\ \forall \bar{t} > 0 \vee \xi(\rho, \bar{t}) = (\bar{r}, \bar{\phi}, 1)\ \forall \bar{t} > 0$, then it converges to $(1,1,1)$, $\lim_{t \to \infty} \xi(\rho, t) = (1,1,1)$. Alternatively, the system must alternate between behavioral equilbria 1 and ϕ. For any $\xi(\rho, 0) = (r, \phi, 1)$, $\dot{r} \geq 0 \wedge \dot{\phi} \geq 0$. Hence, we move towards $(1,1,1)$. For any $\xi(\rho, 0) = (r, \phi, \phi)$, $\dot{\phi} > 0$ and the system as a whole approaches $(1,1,1)$. Hence, the dynamic system moves towards $(1,1,1)$ also if changes in behavioral equilibira are to occur. □

Proof of lemma 5.2. $s(1,1) > c(1)$ implies there is a neighborhood U of $(1,1,1)$ s.t. $s(\phi,\psi) > c(r)\ \forall \rho = (r,\phi,\psi) \in U$. Hence, $\xi(\rho,0) = (r,\phi,1) \forall \rho = (r,\phi,\psi) \in U$. "If": Suppose $\gamma > 0$. $\psi^* = 1 \Rightarrow \dot{r} \geq 0$. $(\psi^* = 1 \wedge r < 1) \Rightarrow \dot{r} > 0$. $\psi^* = 1 \Rightarrow \dot{\phi} = \phi(1-\phi)\gamma > 0$. For all $\rho = (r,\phi,\psi) \in U$, $\dot{\phi} > 0$ if $\phi < 1$ and $\dot{\phi} = 0$ if $\phi = 1$. Hence, $\xi(\rho,t)$ approaches $(1,1,1)$ at any $\rho \in U \setminus \{(1,1,1)\}$, implying it is asymptotically stable. "Only if": Suppose $\gamma = 0$ and consider some $\rho = (1,\phi,1) \in U$. $r = 1 \Rightarrow \dot{r} = 0$ and $s(\phi,\psi) > c(r) \Rightarrow \dot{\psi}^* = 0$. Moreover, $\gamma = 0 \Rightarrow \dot{\phi} = \phi(1-\phi)\gamma = 0$. Hence, the dynamic system remains at rest and does not return to $(1,1,1)$. □

Proof of proposition 5.3. Follows from $s(\phi,\phi) < c(r) < s(\phi,\phi) + p \Rightarrow \phi \in \Psi^*(r,\phi) \wedge \dot{\psi}^* = \dot{\phi}$, $(\kappa s(\phi,\phi) + \gamma = c(r) \wedge \psi^* = \phi) \Rightarrow \dot{\phi} = 0 = \dot{\psi}^*$, and $(r\delta(r) = e(1-\phi) \wedge \psi^* = \phi) \Rightarrow \dot{r} = 0$. □

Proof of proposition 5.4. Consider any $(r,\phi,\phi) \in (0,1]^3$ of proposition 5.3. Note that for all $\hat{\rho} = (\hat{r},\hat{\phi},\hat{\psi})$ in some neighborhood U of (r,ϕ,ϕ), $s(\hat{\phi},\hat{\psi}) < c(\hat{r}) < s(\hat{\phi},\hat{\psi}) + p$. Hence, $\xi(\hat{\rho},0) = (\hat{r},\hat{\phi},\hat{\phi})\ \forall \hat{\rho} \in U$. Since $\hat{\psi}^* = \hat{\phi}\ \forall \hat{\rho} \in U$, we can restrict our attention to the reduced dynamic system consisting of (1) $\dot{\hat{r}} = \hat{r}\delta(\hat{r}) - e(1-\hat{\phi})$ and (2) $\dot{\hat{\phi}} = v\hat{\phi}(1-\hat{\phi})(\kappa\,(s'_\phi(\phi,\phi) + s'_\psi(\phi,\phi)) + \gamma - c(\hat{r}))$ to analyze asymptotic stability of (r,ϕ,ϕ). Linearization of this reduced system around equilibrium point (r,ϕ,ϕ) yields the Jacobian matrix $J = \begin{pmatrix} \delta(r)+r\delta'(r) & e'(1-\phi) \\ -v\phi(1-\phi)c'(r) & v\phi(1-\phi)\kappa\,(s'_\phi(\phi,\phi)+s'_\psi(\phi,\phi)) \end{pmatrix}$. The equilibrium point is asymptotically stable if trace $tr(J) < 0$ and determinant $det(J) > 0$. $tr(J) = (r\Delta'(r) + \delta(r)) + v\phi(1-\phi)\kappa\,(s'_\phi(\phi,\phi) + s'_\psi(\phi,\phi)) < 0$ if the second condition of proposition 5.4 holds and $det(J) = (\delta(r) + r\delta'(r)) * v\phi(1-\phi)\kappa\,(s'_\phi(\phi,\phi) + s'_\psi(\phi,\phi)) + e'(1-\phi) * v\phi(1-\phi)c'(r) > 0$ if the first condition of proposition 5.4 holds. □

Proof of proposition 5.5. Let $\check{r} < \hat{r}$ be s.t. $c(\check{r}) = s(1,0) + p < s(1,1) + p = c(\hat{r})$. Such $\check{r},\hat{r}$ exist due to continuity of c. $(r,\phi,\psi^*) \in cl(\Omega_{BLC}) \Rightarrow \phi = 1 \Rightarrow \dot{\phi} = 0$. For all $\rho \in \{(r,1,1)\}_{\check{r} \leq r < \hat{r}}$, $(\psi^* = 1 \wedge r < 1) \Rightarrow \dot{r} > 0$ and $s(1,1) + p > c(r) \Rightarrow \psi^* = 1 \wedge (\dot{\psi}^* \in \{0,\dot{\phi}\} \Rightarrow \dot{\psi}^* = 0)$. Hence, $\dot{\phi} = 0$, $\dot{\psi}^* = 0$, and $\dot{r} > 0$. r approaches $\hat{r}$. However, $\xi((\hat{r},1,1),0) \neq (\hat{r},1,1)$ since $s(1,1) + p = c(\hat{r}) \Rightarrow 1 \notin \Psi^*(\hat{r},1)$. Since $\phi = 1$, $\xi((\hat{r},1,1),0) = (\hat{r},1,0)$. For all $\rho \in \{(r,1,1)\}_{\check{r} < r \leq \hat{r}}$, $(\psi^* = 0 \wedge r > 0) \Rightarrow \dot{r} < 0$ and $s(1,0) + p < c(r) \Rightarrow \psi^* = 0 \wedge (\dot{\psi}^* \in \{0,\dot{\phi}\} \Rightarrow \dot{\psi}^* = 0)$. Hence, $\dot{\phi} = 0$, $\dot{\psi}^* = 0$, and $\dot{r} < 0$. r approaches $\check{r}$. However, $\xi((\hat{r},1,0),0) \neq (\hat{r},1,0)$ since $s(1,0) + p = c(\check{r}) \Rightarrow 0 \notin \Psi^*(\hat{r},1)$. Since $\phi = 1$, $\xi((\hat{r},1,0),0) = (\hat{r},1,1)$. The above describes the cyclic movement across Ω_{BLC}, implying the dynamic system reaches each $\rho \in \Omega_{BLC}$ and always remains in Ω_{BLC}. □

Sketch of the proof of proposition 5.6. Consider any Ω_{BLC} of proposition 5.5 and let r_{max} and r_{min} be the maximum and minimum resource level. We begin by showing the first condition. For a given γ, let $\kappa > \min\{\frac{s(1,1)+p-\gamma}{s(1,1)}, 0\}$. Consider some neighborhood U of Ω_{BLC} s.t. $\forall \rho \in U$, $\xi(\rho,0) = (r,\phi,\phi) \Rightarrow \dot{\phi} > 0$. Such a U must exist, since $\kappa > \min\{\frac{s(1,1)+p-\gamma}{s(1,1)}, 0\} \Rightarrow \kappa s(1,1) + \gamma > s(1,1) + p$ and $\xi(\rho,0) = (r,\phi,\phi) \Rightarrow s(1,1) + p > c(r)$. Moreover, if $s(1,1) < p + s(1,0)$, let U be s.t. $\xi(\rho,0) \neq (r,\phi,1)\ \forall \rho \in U$. If $s(1,1) < p + s(1,0)$, such U exists since (1) $\xi(\rho,0) = (r,\phi,1) \Rightarrow s(\phi,1) > c(r) \Rightarrow s(1,1) > c(r)$, but $p + s(1,0) \leq c(r)\ \forall \rho \in \Omega_{BLC}$, and (2) all involved functions are continuous. Moreover, let U be s.t. $\forall \rho \in U$, $\delta(r) * r > e(1-\phi)$. Such a neighborhood must exist since (a) $\phi = 1\ \forall \rho \in \Omega_{BLC}$, (b) $\delta(r) * r > e(0)$, and (c) continuity of all involved functions. Hence, $\xi(\rho,0) = (r,\phi,\phi) \Rightarrow s(1,1) + p > c(r)$. In conjunction, the above implies that $\forall \rho \in U$, (1) $\xi(\rho,0) = (r,\phi,0) \Rightarrow \dot{\phi} \geq 0 \wedge \dot{r} < 0$, (2) $\xi(\rho,0) = (r,\phi,\phi) \Rightarrow \dot{\phi} > 0 \wedge \dot{r} > 0$, and (3) $\xi(\rho,0) = (r,\phi,1) \Rightarrow \dot{\phi} \geq 0 \wedge \dot{r} > 0$. Lastly, let U be s.t. $\forall \rho \in U$, (1) $s(\phi,\phi) + p > s(\phi,1)$ and (2) $s(\phi,0) + p > c(0)$. Such U exists since (1) $p > 0$ and $\phi = 1$ $\forall \rho \in \Omega_{BLC}$ and (2) $s(1,0) + p > c(0)$ and continuity of $s(\phi,0)$. Note that since $\dot{\phi}$ follows the replicator dynamics, $\nexists t > 0$ s.t. $\xi(\bar{\rho},t) = (r,1,\psi^*)$ for any $\bar{\rho} \in U/\Omega_{BLC}$.

Consider any $U_2 \subset U$ s.t. $(y,x,0) \in U$, where $x = \arg\min_{\phi}\{(r,\phi,\psi) \in U\}$ and y solves $s(x,0) + p = c(y)$. Consider any $\rho \in U_2/\Omega_{BLC}$. If $\xi(\rho,0) = (r,\phi,0)$, $\dot{\phi} \geq 0 \wedge \dot{r} < 0$. The resource stock diminishes, and the social norm (weakly) increases but never reaches one. Let $\bar{t}$ be s.t. $\xi(\rho,\bar{t}) \neq (\bar{r},\bar{\phi},0)$ and $\xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},0) \in U$ $\forall \hat{t} < \bar{t}$. Hence, at $\bar{t}$ it holds that $s(\bar{\phi},0)+p = c(\bar{r})$. Note that such a $\bar{t} > 0$ exists, since $s(\phi,0)+p > c(0)$ and $\hat{\phi} \geq \phi$ $\forall \xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},0), \hat{t} < \bar{t}$. Moreover, $\xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},0) \in U$ since $\hat{\phi} > \phi$ and $\hat{\phi} \geq \phi \geq x \Rightarrow \hat{r} \geq y$. By the same reasoning, $\xi(\rho,\bar{t}) \in U$. The solution eventually reaches a point in U, where full exploitation is no longer a behavioral equilibrium. Behavior transits to partial or no exploitation. Next, consider any $\rho \in U/\Omega_{BLC}$. If $\xi(\rho,0) = (r,\phi,1)$, $\dot{\phi} \geq 0 \wedge \dot{r} > 0$. The resource stock increases, and the social norm (weakly) increases but never reaches one. Let $\bar{t}$ be s.t. $\xi(\rho,\bar{t}) \neq (\bar{r},\bar{\phi},1)$ and $\xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},1) \in U$ $\forall \hat{t} < \bar{t}$. Note that such a $\bar{t} > 0$ exists, since $s(\hat{\phi},0) < s(1,1)+p < c(r_{max})$. $\xi(\rho,\bar{t}) \neq (\bar{r},\bar{\phi},1) \Rightarrow s(\bar{\phi},1) = c(\bar{r}) \Rightarrow s(\bar{\phi},\bar{\phi})+p > c(\bar{r})$. Moreover, $(\lim_{t\to\bar{t}} \xi(\rho,t) = (\bar{r},\bar{\phi},1) \wedge s(\bar{\phi},\bar{\phi}) + p > c(\bar{r})) \Rightarrow \xi(\rho,\bar{t}) = (\bar{r},\bar{\phi},\bar{\phi})$. $\xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},1) \in U$ $\forall \hat{t} < \bar{t}$ holds since $\hat{\phi} \geq \phi$ and $\hat{r} \in (r,r_{max})$. By analog reasoning, $\xi(\rho,\bar{t}) \in U$. The solution eventually reaches a point in U, where no exploitation is no longer a behavioral equilibrium and transits to partial exploitation. If $\xi(\rho,0) = (r,\phi,\phi)$, $\dot{\phi} > 0 \wedge \dot{r} > 0$. The resource stock increases, and the social norm increases but never reaches one. Let $\bar{t}$ be s.t. $\xi(\rho,\bar{t}) \neq (\bar{r},\bar{\phi},\bar{\phi})$ and $\xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},\hat{\phi}) \in U$ $\forall \hat{t} < \bar{t}$. Such a $\bar{t} > 0$ exists since $s(1,1) + p < c(r_{max})$. $\xi(\rho,\hat{t}) = (\hat{r},\hat{\phi},\hat{\phi}) \in U$ $\forall \hat{t} < \bar{t}$ holds since $\hat{\phi} > \phi$ and $\hat{r} \in (r,r_{max})$. The solution eventually reaches a point in U, where partial exploitation is no longer a behavioral equilibrium. Behavior transits to full or no exploitation.

The above implies that starting at any $\rho \in U_2$, we remain in U and experience behavioral transitions. Throughout, $\dot{\phi} \geq 0$. At times, however, $\psi^* = \phi$ and, thus, $\dot{\phi} > 0$. The dynamic system may, at times, leave U_2, when $\psi^* = 0$. Since throughout $\dot{\phi} \geq 0$ holds, we, afterwards, transit to partial or full exploitation returning $\xi(\rho,t)$ to U_2. At each time $\bar{t}_1$ s.t. $\xi(\rho,\bar{t}_1) = (\bar{r}_1,\bar{\phi}_1,\bar{\psi}_1^*)$ and $\exists \bar{t}_2 < \bar{t}_1$ s.t. $\xi(\rho,\bar{t}_2) = (\bar{r}_2,\bar{\phi}_2,\bar{\psi}_2^*)$ and $\bar{r}_2 = \bar{r}_1$, it must hold that $\bar{\phi}_2 > \bar{\phi}_1$. The social norm approaches 1. Hence, the outer resource stocks (at which the behavioral transitions away from $\psi^* = 0$ and $\psi^* = 1$ occur) move towards r_{min} and r_{max} respectively. The resource stock where the transition away from $\psi^* = \phi$ lies somewhere in between or society never reaches this behavioral equilibrium (since $s(1,1) < p + s(1,0)$, see above). The system converges to Ω_{BLC}. The proof for the second statement of the proposition mimics the above.

Lastly, we turn to the third statement. Suppose the limit cycle is asymptotically stable. From the above, we know that at any $\rho \in U$, $\xi(\rho,t)$ approaches Ω_{BLC} cyclic. An increase in γ and κ increases $\dot{\phi}$ at any $\rho \in U$. Thus, the limit cycle must also be asymptotically stable. □

Sketch of the proof of proposition 5.7. First, suppose that $\kappa s(1,1) + \gamma > 0$. Let $\bar{p} > 0$ be s.t. $s(1,1) + \bar{p} < \kappa s(1,1) + \gamma$. Consider any $p < \bar{p}$. There is some U of Ω_{BLC} s.t. $s(\phi,\phi) + p < \kappa s(\phi,\phi) + \gamma$ for all $\rho = (r,\phi,\psi) \in U$. Hence, $\forall \rho \in U$, $\xi(\rho,0) = (r,\phi,\phi) \Rightarrow \psi^* = \phi \Rightarrow c(r) <$

$s(\phi,\phi)+p<\kappa s(\phi,\phi)+\gamma \Rightarrow \dot{\phi}>0$ and $\xi(\rho,0)\neq(r,\phi,\phi)\Rightarrow \psi^*\in\{0,1\}\Rightarrow\dot{\phi}\geq 0$. We move towards the perfect social norm. By similar reason as in the sketch of proof 5.6, the resource evolves cyclic, while behavior transits between different equilibria, and the solution converges to Ω_{BLC}.

Second, suppose $\gamma>0$ and consider some $\epsilon>0$. Suppose for a moment that $p=0$. Consider $\check{\rho}\in\{\rho\in[0,1]^3 : s(\phi,1)=c(r)\}$ s.t. for $\hat{\rho}\in\{\rho\in[0,1]^3 : s(\phi,0)=c(r)\}$ satisfying $\lim_{t\to\bar{t}}\xi(\hat{\rho},t)=\check{\rho}$ for $\bar{t}>0$ and $\xi(\hat{\rho},t)=(r,\phi,0)$ for all $t\in[0,\bar{t})$ it holds that $|x-\hat{\rho}|<\epsilon$ for some $x\in cl(\Omega_{BLC})$. Let $\delta=|x-\check{\rho}|$ for some $x\in cl(\Omega_{BLC})$. Consider an arbitrary $\hat{p}>0$ and $0<p<\hat{p}$. Note that for any ϕ, $s(\phi,0)+p=c(r)$ is solved for lower values of r than $s(\phi,0)+\hat{p}=c(r)$. Consider $\tilde{\rho}\in\{\rho\in[0,1]^3 : s(\phi,0)+\hat{p}=c(r)\}$ satisfying $\lim_{t\to\bar{t}}\xi(\tilde{\rho},t)=\check{\rho}$ for $\bar{t}>0$ and $\xi(\tilde{\rho},t)=(r,\phi,0)$ for all $t\in[0,\bar{t})$.

Let $d_1=|\check{\phi}-\hat{\phi}|$. Consider the dynamic system at $\check{\rho}$ and note that $\psi^*=\phi$. Look at the first point in this two dynamic system, where we again reach social norm $\hat{\phi}$. Let p be so small, that that this point it holds $c(r)>s(\hat{\phi},\hat{\phi})+p$. Hence, before we reach this point, society jumps into the equilibrium of full exploitation. $\xi(\rho,t)$ experiences a resource decrease but an norm increase due to institutional pressure. Eventually, $\xi(\rho,t)$ reaches the point where behavior transits to $\psi^*=1$. After that, the solution experiences a norm increase through institutional pressure and a resource increase. Eventually, $\xi(\check{\rho},t)$ reaches some $\bar{\rho}_1$ at $\bar{t}_1$ s.t. $\xi(\check{\rho},\bar{t}_1)=(\bar{r}_1,\bar{\phi}_1,\bar{\psi}_1^*)$ and $\exists\bar{t}_2<\bar{t}_1$ s.t. $\xi(\check{\rho},\bar{t}_2)=(\bar{r}_2,\bar{\phi}_2,\bar{\psi}_2^*)$ and $\bar{r}_2=\bar{r}_1$. It must hold that $\bar{\phi}_2>\bar{\phi}_1$. The system has moved towards the limit cycle.

Lastly, we can show that for any $\rho\in U$ s.t. $\delta>|x-\rho|$, $\xi(\rho,t)$ also approaches the limit cycle. The underlying reasoning is that at any such ρ, the solutions on the two-dimensional systems of ϕ and r for a given ψ^* cannot intersect with those of $\xi(\check{\rho},t)$. Hence, the solution $\xi(\rho,t)$ cannot intersect with $\xi(\check{\rho},t)$ at any t. Moreover, the norm only decreases and, thus, $\xi(\check{\rho},t)$ distances from Ω_{BLC} if $\psi^*=\phi$. However, at any such $\xi(\check{\rho},\check{t})=(\check{r},\check{\phi},\check{\phi})$, we can show that there is t s.t. $\xi(\rho,t)=(\check{r},\phi,\phi)$ where $\phi<\check{\phi}$. From this, we can follow that $\xi(\check{\rho},\check{t})$ also moves towards Ω_{BLC}. $\square$

Proof of lemma 5.3. For condition 1a, consider some N of 1 as described. Consider some U of $\check{\Omega}_{BLC}\cup\hat{\Omega}_{BLC}$ s.t. $\phi\in N\ \forall\rho\in U$. Consider some $\rho\in U$ s.t. $\check{\xi}(\rho,0)=(r,\phi,\phi)\wedge\dot{\phi}>0$ at $(\check{s},\check{c})$. $\dot{\phi}>0\Rightarrow\kappa\check{s}(\phi,\phi)+\gamma>\check{c}(r)\Rightarrow\kappa\hat{s}(\phi,\phi)+\gamma>\hat{c}(r)$. To see why this holds, consider $\bar{r}$ s.t. $\kappa\check{s}(\phi,\phi)+\gamma=\check{c}(\bar{r})$. Since $[\hat{c}^{-1}](\kappa\hat{s}(\phi,\phi)+\gamma)=[\check{c}^{-1}](\kappa\check{s}(\phi,\phi)+\gamma)$, it follows $\kappa\hat{s}(\phi,\phi)+\gamma=\hat{c}(\bar{r})$. $\kappa\check{s}(\phi,\phi)+\gamma>\check{c}(r)\Rightarrow r>\bar{r}\Rightarrow\kappa\hat{s}(\phi,\phi)+\gamma>\hat{c}(r)$. Moreover, $\kappa\hat{s}(\phi,\phi)+\gamma>\hat{c}(r)\Rightarrow\hat{s}(\phi,\phi)+\gamma>\hat{c}(r)$. $\hat{\xi}(\rho,0)=(r,\phi,\phi)$ at $(\check{s},\check{c})\Rightarrow\check{s}(\phi,\phi)+p>\check{c}(r)>\check{s}(\phi,\phi)\Rightarrow\hat{s}(\phi,\phi)+p>\hat{c}(r)$. Combining these insights, yields $\hat{s}(1,1)<\hat{c}(r)<\hat{s}(1,1)+p$ and $\kappa\hat{s}(1,1)+\gamma>\hat{c}(r)$, which imply $\hat{\xi}(\rho,0)=(r,\phi,\phi)\wedge\dot{\phi}>0$ at $\hat{s},\hat{c}$. The proof for condition 1b is analog. Condition 2 follows straightaway from how $\dot{\phi}$ depends on $s(\phi,\psi)$ and $c(r)$. $\square$