

The B.E. Journal of Economic Analysis & Policy

Topics

Volume 10, Issue 1

2010

Article 13

Identities in the Commons: The Dynamics of Norms and Social Capital

Erwin Bulte*

Richard D. Horan†

*Wageningen University & Tilburg University, erwin.bulte@wur.nl

†Michigan State University, horan@msu.edu

Recommended Citation

Erwin Bulte and Richard D. Horan (2010) "Identities in the Commons: The Dynamics of Norms and Social Capital," *The B.E. Journal of Economic Analysis & Policy*: Vol. 10: Iss. 1 (Topics), Article 13.

Identities in the Commons: The Dynamics of Norms and Social Capital*

Erwin Bulte and Richard D. Horan

Abstract

This paper provides a formal analysis of the evolution of cooperation in the management of common property resources. We develop a dynamic model that includes moral norms or a sense of ‘identity,’ and show that cooperation may – but need not – be an equilibrium outcome in the absence of intervention by a managing agency or punishment by peers. We demonstrate that outside intervention has ambiguous effects when identity matters – it may reduce welfare of the agents harvesting the stock.

KEYWORDS: internalization of moral norms, changing preferences, social capital, regulating common pools

*We would like to thank two anonymous referees for helpful comments and suggestions. Remaining errors are our own. Erwin Bulte gratefully acknowledges financial support from the Dutch Organisation for Scientific Research, N.W.O. (grant nr 452-04-333).

1. INTRODUCTION

Property rights for many natural and environmental resources are imperfectly defined or enforced. While imperfect property rights in the presence of appropriation externalities may result in excessive use, circumstances also arise where communities are able to overcome such social dilemmas. Various terms, including social capital, community governance and social cohesion, are used to capture the idea that trust and a shared willingness to live by norms are key factors determining how efficiently common property resources (CPR) are used. In recent years academic interest in these issues is flourishing.

The economic literature on CPR management has focused mainly on two interrelated topics. First is the efficiency and distributional consequences of CPR management versus alternative forms of property rights [Weitzman, 1974, de Meza and Gould, 1987, Brito et al., 1997; Baland and Francois, 2005]. Second, a rapidly growing literature explores the motivations for cooperation in the commons – the ‘social capital’ focus mentioned above. Why do people forego private gains for the public good? This question strikes at the heart of the overlapping research agendas of institutional, environmental and development economics, and it is the main issue addressed in this paper.

Economists argue that agents will cooperate when their best interest is to do so. Among the potential mechanisms to align social and private incentives are reputation effects, monitoring and punishment by peers, and future gains from cooperation [Ostrom, 1990, Ostrom et al., 1992, 1994, Sethi and Somanathan, 1996, Mailath and Samuelson 2006]. The difficulty, however, has been to reconcile the theoretical behavior of *Homo economicus* with the plethora of behaviors observed in ‘the field’ and in laboratory settings. Three complementary traditions in economics attempt to do so. First, cooperation can be sustained as a sub-game perfect equilibrium in the presence of punishment opportunities – including ostracism – when forward-looking agents interact in (infinitely) repeated and dynamic games.¹ However, the Folk Theorem indicates that many other stable equilibria may emerge as well, and generally does not indicate which equilibrium will materialize. Second, a few papers analyze the accumulation of trust through repeated interaction and improved information, facilitating cooperative behavior over time [Watson, 1999, Fafchamps 2004]. Finally, cooperation may emerge as a stable Nash equilibrium in evolutionary models involving interactions across individuals, in which myopic agents replicate the behavior of more successful individuals in the community [Sethi and Somanathan, 1996, Osés-Eraso and Viladrich-Grau, 2007].

¹ E.g., see Kandori [1992], and Kranton [1996], on sustaining cooperative relationships; for a focus on common property resources refer to Dutta [1995], Kranton [1996], Hannesson [1997], Polasky et al. [2006], and Tarui [2007].

Our model fits into the third tradition. Prior work within this tradition has analyzed outcomes under different types of strategic interactions – that is, under different assumptions about how individuals’ conservation and regulatory actions affect others’ utility. Sethi and Somanathan model two types of interactions. The first is a common harvest externality associated with the common pool nature of the resource. The second interaction is based on the assumption that agents who choose low effort can punish individuals choosing high harvest efforts, though at some cost to themselves. This increases the basin of attraction associated with the low effort equilibrium. However, multiple equilibria are still possible, with each potential stable equilibrium involving a homogenous population: all agents either choose high effort or they all choose low effort.

Osés-Eraso and Viladrich-Grau [hereafter, OEVG] adopt a similar harvest model but they do not model punishment. Rather, they assume that, beyond harvest returns, choosing low effort leads to inclusion in a social group that yields benefits, such as respect and friendship, to its members. These benefits are increasing in the proportion of individuals choosing low effort. This effect counteracts the strategic complementarities that stem from the harvest relation and, in contrast to Sethi and Somanathan, can result in stable equilibria involving heterogeneous populations, with some agents choosing low effort and others choosing high effort. This outcome is consistent with observations of reality, as compliant and non-compliant behaviors have co-existed for long stretches in many traditional societies.

We extend this literature by investigating a broader set of interactions that combines features of the earlier models. As with prior work, payoffs depend on effort choices. Additionally, we model choices about personal values that influence the payoffs individuals derive from their own effort choices as well as from social interactions. Values are therefore not simply based on resource use decisions, and social interactions can occur and produce benefits in spite of disagreements over these decisions.

In our model, personal values are based on the notion that people have an intrinsic motivation to behave cooperatively by following “moral” or “social norms” [Dowell et al. 1998, Lindbeck 1997].² The idea that following social

² Norms are expectations and beliefs about how one should behave. Moral norms are internalized rules of conduct, and therefore potentially self-policing [Baland and Platteau, 1996]. Violating the norm is intrinsically costly as it causes feelings of guilt and shame. The adverse consequences associated with norm violating are thus not necessarily punishment by peers or regulators. Much of the literature on cooperation and social capital emphasizes the key role of enforcement [e.g., McCarthy et al., 2001] and willingness to ‘punish’ those who do not live by the norm [e.g., Bowles and Gintis, 2002]. While norms (and the emotions associated with norm violations by others) facilitate punishment, we will abstract from this in what follows. The paper’s main results may apply in other contexts than management of common pools, but the model and discussion are

norms may be privately optimal when norm-violation causes disutility goes back to Becker [1976], and has recently been formalized and extended by Akerlof and Kranton [2000; hereafter AK]. In particular, AK introduce social psychology's concept of '*identity*' by augmenting utility to account for 'identity gains' when social norms are followed, and 'identity losses' when norms are violated.³ Different individuals may live by different norms, experiencing disutility when deviating from them. Moreover, individuals may acquire utility from associating with others like themselves.

Importantly, our model allows people to update their identities, and hence change the costs and benefits to which they are exposed.⁴ People have the ability to adapt their psychology (i.e., their 'beliefs' about true norms) to reduce dissonance between behavior and beliefs, or to associate with another group within the community. While the AK model allows for two mechanisms to "maintain a sense of unity" between prescriptions and norms (that is; via actions to reduce anxiety and via cognitive dissonance), we essentially introduce a third mechanism by allowing individuals to adjust their identity over time. In our model, resource stocks, identities, and behaviors co-evolve over time. We therefore endogenize not only the distribution of behavior and payoffs, but also the distribution of payoff functions.

What evidence motivates the (dynamic) identity model? Economic experiments support the idea that there are different 'types' within populations – some more inclined to cooperation and others more inclined to defection [e.g., Fehr and Gächter, 2000, Offerman et al 2006, Fischbacher et al. 2001]. Similar observations emerge from case studies on CPR management [e.g. Feeny et al., 1996, Baland and Platteau, 1996, Henrich et al., 2001]. But such moral imperatives are not written in stone. For economic models of contingent moral motivation, refer to Brekke et al. [2003] and Nyborg et al. [2006]. Focusing on the case of CPR management, Baland and Platteau [1996, 123] write that "moral norms are subject to erosion; they form a 'social capital' ... liable to depreciation, especially so if norm-abiding individuals come to realize that many people around them behave opportunistically." When surrounded by 'defectors,' cooperators start discounting the relevance of living by the norms and, *vice versa*, defectors

cast in terms of CPRs as this is an application where moral norms obviously matter, and because this application is readily amenable for a dynamic analysis.

³ This approach is distinct from work focusing on "external" value of having a certain identity – think of reputation effects in repeated games [e.g. Nowak 2006].

⁴ Our model is of course linked to existing work on the endogenous evolution of preferences. Kuran and Sandholm (2008) study socialization and self-persuasion in the context of cultural integration, Bisin and Verdier (2001) and Francois and Zabojnik (2007) analyse parental socialization and economic outcomes. Related papers in a parallel literature rooted in evolutionary game theory include Guth and Yaari (1992) and Dekel et al. (2007).

may internalize norms when it is in their interest to do so.⁵ Sociological experiments [Taifel, 1978, Haslam 2001] confirm the malleability of group identification and support the idea that identities can change over time.

This paper has four main contributions. First, we extend the identity model by making it dynamic, modeling how the abundance of ‘types with a certain payoff function’ in the community evolves in response to payoff differences. This enables us to analyze the accumulation or erosion of social capital, as measured by the distribution of beliefs facilitating or hampering conservation (in contrast to OEVG, who measure social capital by the number of people *behaving* as a conservationist). This approach, whereby individuals self-select into a certain identity in response to interactions with others and with nature, provides a new mechanism to understand the evolution of preferences in individuals and populations.

Second, we provide a novel theoretical underpinning for the observation that cooperation in the commons can – but need not – occur in the absence of regulation by outsiders or peers. This is consistent with the observation that some CPRs are managed sustainably whereas others suffer from the ‘tragedy of the commons’ [e.g. Benin and Pender, 2006, Lopez, 1998, Ahuja, 1998].

Third, our model produces a variety of potential equilibria, encompassing predictions from earlier theoretical models as special cases. Unlike Sethi and Somanathan but consistent with OEVG, we find mixed outcomes with only some individuals conserving might emerge as an equilibrium outcome. However, and unlike OEVG, we also identify cases where heterogeneous stable equilibria cannot emerge.

Fourth, we show that introducing effort-based incentives to encourage conserving behavior yields ambiguous effects in a setting where identity matters. Specifically, an incentive that encourages conserving *behavior* (low effort) by all may not encourage everyone to switch to a conservationist *identity*. As a result, the long-run welfare effects of the regulation are ambiguous. A regulation that induces everyone to adopt conserving behavior but not a conservationist identity will result in a welfare loss relative to the case of a homogeneous population of conservationists. In general, for policies to be effective or efficient, regulators need to know the underlying distribution of identities in the population of individuals that they try to regulate.

The paper is organized as follows. In section 2 we introduce the ‘identity’ concept as well as identity-dependent utility gains and losses. In section 3 we develop a dynamic model where individuals can both change their behavior as

⁵ Interestingly, Baland and Platteau [1996, 122] mention evidence that people are more responsive to the costs of living by moral norms than by the potential gains that can be had by discounting them (in our parlance; adopting another identity). In what follows we ignore this asymmetry between benefits and costs, and simply consider overall payoff differences.

well as their beliefs (but accounting for the fact that changing behavior is easier than changing identity). In section 4 we analyze the model and distinguish between different scenarios. In section 5 we introduce regulation. The conclusions ensue.

2. IDENTITY IN COMMON POOL MANAGEMENT: A STATIC MODEL

Society has different social categories (types or identities), with different forms of “prescribed behavior.” A simple example is between ‘men’ and ‘women.’ Different ideal behaviors are usually associated with these identities, and “violating the prescribed behavior causes anxiety and discomfort in oneself and in others. Gender identity, then, changes the payoffs from different actions” [AK, p. 716].⁶ While the ‘gender’ identity is usually pre-selected by nature, people often can choose their identity and then experience significant peer pressure to behave in a particular manner. Examples include joining the army, student fraternities, or a street gang. Rules of behavior are associated with each of these social categories, and violating these rules causes anxiety that individuals act to reduce.

We adopt the basic AK model and modify it for the CPR case. Two basic assumptions are that people have identity-based payoffs derived from: (i) their own actions; and (ii) from others’ actions. We employ these assumptions in a CPR model with two identities: conservationists and non-conservationists. The latter are also labeled as “defectors” in what follows, although the individuals in question probably prefer the label “entrepreneurs” for themselves. We begin with a static model, but later make it dynamic, allowing agents to update their identity. This extension is obviously interesting and relevant in the context of evolving moral norms, and the accumulation and erosion of social capital.

Each identity-type has its own social or moral norm defining the preferred level of harvesting effort (henceforth labeled effort). For simplicity, but without great loss, we follow Sethi and Somanathan [1996] and OEVG [2007] and consider the tractable case of two fixed effort levels: high and low effort. The moral norm of conservationists is to choose low harvest effort level E_L , which represents cooperative behavior. In the words of OEVG [2007: 396]: “through years of communal living, society members have developed a sense of what level of exploitation is ‘desirable,’ or have reached an understanding as to what effort level will permit a reasonable exploitation of the resource.” This may involve respecting certain restrictions (perhaps taboos) with respect to the timing of harvesting or the types of technologies that are allowed. Defectors’ preferred

⁶ Note that this example also nicely illustrates the importance of treating ‘identity’ as a dynamic concept, subject to change. For example, the mainstream norm regarding whether ‘decent women’ should engage in paid labor has changed considerably during recent decades (e.g. Vendrik 1993).

behavior is to exploit short-term opportunities for personal gain, even if doing so imposes costs on others. For simplicity, the norm for defectors is to ‘defect’ from the cooperation strategy by choosing the highest feasible effort level E_H , so that $E_H > E_L$.

We begin with a general model to introduce some basic concepts. In the following section, we present a fully-specified model that is more amenable to detailed analysis. Consider a community with a large fixed number of individuals. The utility of an individual of identity type I ($I = C$ for conservationists; $I = D$ for defectors) who chooses behavior J ($J=L$ for low effort and $J=H$ for high effort) is denoted

$$(1) \quad U_{IJ}(x, z, S).$$

In (1), x is the overall proportion of individuals with conservationist identity in the community. The variable z is the overall proportion of individuals who cooperate (i.e., respect the norm of conservationists) by choosing low effort. Specifically, $z = xz_C + (1-x)z_D$, where z_I is the proportion of identity-type I individuals who cooperate (note that $z_D \neq 1 - z_C$) – as will become evident below both defectors and cooperators may chose to select low effort levels. Finally, S is the resource stock. Individuals are myopic and take S as given, though later we show how the stock changes over time in response to individuals’ harvest choices.

One of the possible characteristics of an identity-based model, as identified by AK, is that individuals may gain utility from being among their own kind (or, alternatively, a utility loss from being with those of other identities):

$$(2) \quad \frac{\partial U_{CJ}}{\partial x} > 0; \quad \frac{\partial U_{DJ}}{\partial x} < 0 \Rightarrow \frac{\partial [U_{CJ} - U_{DJ}]}{\partial x} > 0.$$

We refer to the final relation in (2) as a *fraternal complementarity*, which indicates the relative welfare of conservationists (defectors) is increasing (decreasing) in the share of cooperators in the community. This part of the model is similar to the benefits of group membership in OEVG’s [2007] model, which is the key feature of their model. The difference is that these benefits depend on identities in our model, whereas they depend on behaviors in OEVG’s model. Moreover, utility of defectors is directly affected by changes in the composition of the population in terms of identities.

In addition to fraternal complementarities, identity-based gains or losses are associated with own behavior and behavior of others. Following AK, one’s actions cause disutility to oneself if these actions violate the prescriptions of one’s own identity. This means violating the norm associated with one’s identity. For a

conservationist, violating the norm of restraint produces ambiguous net welfare effects: it raises harvest levels but also involves psychological costs so that the net effect is ambiguous *a priori*. In contrast, norm violation unambiguously produces welfare losses for a defector: harvest revenues are reduced and there is also an identity cost.

In addition, one's actions cause disutility to others (external costs) if these actions violate the norms associated with others' identities. That is, others' actions that run counter to one's own norms cause anxiety: $\partial U_{CJ} / \partial(1-z) < 0$ and $\partial U_{DJ} / \partial z < 0$. We refer to these costs as outrage costs, and in the specified model below we show that outrage costs can generate *strategic complementarities*: $\partial[U_{CJ} - U_{DJ}] / \partial z > 0$. The relative benefits from being a conservationist are larger (smaller) if others cooperate (defect); alternatively, it also means that the relative benefits from being a defector are smaller (larger) when others cooperate (defect). Though different in spirit, these external costs are mathematically analogous to the punishment and enforcement costs of Sethi and Somanathan's [1996] model, which is the key feature of their model.

2.1 A specified model

Our specification of the common pool model is analogous to Sethi and Somanathan [1996] and OEVG [2007] in that we define utility simply as the expected payoffs to individuals, which is the sum of harvest profits and the additional costs and benefits associated with identities and norms. We also follow their assumptions of a fixed price for harvests, a fixed unit cost of effort, and that harvests depend on both the level of effort employed and also the resource stock level. For simplicity, we normalize the harvest price to unity and the effort cost to zero, so that harvests equal profits. An individual's harvests are defined by a scaled Schaefer harvest function, $E_i S$ where $i = L, H$ indexes 'low' and 'high' effort levels, respectively. Note that this harvest function explains why utility in (1) depends on S . Assume that 'low' harvest effort E_L results in larger long-run equilibrium profits than does 'high' harvest effort E_H . If some individuals choose 'high' harvest effort E_H , then sustainable harvest falls, but of course short-run profits of these defectors rises. Hence, a social dilemma eventuates.

Our only substantive departures from Sethi and Somanathan [1996] and OEVG [2007] are that (i) those models assume diminishing returns to individual effort due to congestion externalities, and (ii) E_L and E_H are stock-dependent in those models but not in ours. If we think of myopic profit maximization as the base case from which each of these models (including ours) is compared, then the effect of congestion externalities is to create the potential for multiple equilibria: everyone either chooses high effort or low effort. By ignoring congestion

externalities, our profit maximization base case exhibits a single equilibrium in which the basin of attraction for defection is larger. This simplifies the analysis and helps us to focus on the identity based benefits and external behavioral costs, which are the key features in our model. Our assumption of stock-independent effort levels is also a simplification, the primary impact of which will be to reduce the number of potential equilibrium stock levels – but not the important qualitative result that more effort generally results in smaller stocks.

Now consider the additional costs and benefits arising from individuals' choices of behaviors and the distribution of identities among resource users. First, consider the *external* impacts of identity, associated with the fraternal complementarity. AK develop one model where individuals suffer an identity loss when encountering individuals with a different identity.⁷ Without any loss we turn this argument around – people don't suffer an identity loss when they meet somebody with a different identity, but instead earn a benefit when meeting somebody of their own type. Liberals like to hang out with liberals because they share some deep values (even if the other occasionally votes conservative). We model this benefit as w_o^I (where, as above, $I = C, D$, and the subscript 'o' refers to "others") so the external benefits for conservationists are simply xw_o^C (and for defectors are $(1-x)w_o^D$).

Next, consider welfare effects when one's own actions conflict with internalized norms. If conservationists choose E_H (which feels like cheating to them, and triggers guilt) they suffer a utility loss, monetized as a cost w_s^C . The superscript 'C' refers to the conservationist identity and the subscript 's' refers to 'self' to indicate that the losses are caused by own behavior. In contrast, if defectors choose E_L (which makes them feel like a wimp foregoing a profitable opportunity), they suffer a utility loss valued at w_s^D .

In addition, we follow AK and introduce cross-strategy impacts – the impacts of others' harvesting strategies on one's utility. Such impacts could arise in various ways, but perhaps the most compelling impacts are negative ones that stem from the inherent tension between cooperation and defection. We focus on these. Individuals who choose high effort, E_H , impose a (normalized) cost k on cooperating conservationists who suffer a utility loss from seeing some of their fellow villagers violating the conservationists' norm.⁸ For instance, they may feel

⁷ We assume no uncertainty about the identity of others. People choose their type and immediately signal it to the rest. The model could be expanded to allow for asymmetric information about identities, which is left for future work.

⁸ This external effect may be a function of the stock, and without loss we could also rewrite the model so that the externality both cooperating conservationists and defectors (as opposed to just cooperating conservationists). This is ignored to streamline the presentation.

angry when others refuse to play by their rules. Similarly, defectors who cooperate suffer a utility loss if they observe others defecting. One reason may be that such observations add salience to the regret that they have missed out on a profitable opportunity, making them feel like a sucker. Denote this cost as g . It is also an outrage cost, but one where the defector is outraged with him/herself. We refer to k and g as the “outrage costs” of others choosing high effort, and these outrage costs are the source of strategic complementarities in our model. We reasonably assume that conservationists do not impose similar costs on defectors and that defectors do not impose outrage costs on conservationists who defect.⁹

Given two identities (conservationists and defectors) and two behaviors (high and low effort), we have to consider four different combinations of identity type and behavior. The expected payoffs for these four combinations are presented in Table 1. Assume for now no fines; $F=0$, we return to fines and regulation in section 5. As before, the variable x in Table 1 denotes the share of conservationists in the population, which will be endogenized below. Note that the expected payoffs from choosing high harvest effort levels for a defector always exceed the payoffs from choosing low harvest levels (in the absence of regulation, this is true regardless of outrage costs). Therefore z_D is always at its equilibrium value of zero (defectors always choose high effort), so that $z = xz_C$ denotes the share of the total population choosing low effort.

Table 1. Expected payoffs for individuals of different identities and behaviors in the presence of regulation.^a

Identity/harvest effort level	Expected payoffs
Conservationist/low ($I=C, J=L$)	$U_{CL} = E_L S + x w_o^C - (1 - z)k$
Conservationist/high ($I=C, J=H$)	$U_{CH} = E_H S - w_s^C + x w_o^C - F$
Average Payoff Conservationists	$\bar{U}_C = z_C [E_L S + x w_o^C - (1 - z)k]$ $+ (1 - z_C) [E_H S - w_s^C + x w_o^C - F]$
Defector/low ($I=D, J=L$)	$U_{DL} = E_L S - w_s^D + (1 - x) w_o^D - (1 - z)g$
Defector/high ($I=D, J=H$)	$U_{DH} = E_H S + (1 - x) w_o^D - F$
Average Payoff to Defectors	$\bar{U}_D = z_D [E_L S - w_s^D + (1 - x) w_o^D - (1 - z)g]$ $+ (1 - z_D) [E_H S + (1 - x) w_o^D - F]$

^a Set $k=g=0$ for the case of “no outrage costs”. Set $F=0$ for the case of no regulation.

⁹ We could easily incorporate such costs but, as will become evident, equilibrium never has cooperators who choose high effort while defectors choose low effort. Hence, we ignore this possibility. Of course, depending on the specific (cultural) context alternative specifications that may lead to such outcomes might be feasible, and in those instances these additional costs may have ramifications for the dynamics and equilibria that eventuate.

Payoffs depend on the distribution of identities and behaviors. This payoff structure potentially allows for multiple qualitatively different equilibria (or combinations of identity and behavior) in the community. Without adding more structure to the model, the outcomes may range from $(x=1, z=1)$ to the polar opposite case of $(x=0, z=0)$. Heterogeneous identities and heterogeneous harvest effort choices are also feasible: co-existence of cooperators and defectors (with cooperators choosing low and defectors choosing high effort levels, but also where both types choose high effort, or even where some cooperators choose low and others choose high effort levels). To analyze the stability of these steady states we develop an explicitly dynamic model – one in which strategies, identities, and resource stocks all evolve over time.

3. A DYNAMIC MODEL

We extend the static model by allowing a unit measure of agents to adjust both identity and behavior over time in response to changing conditions, recognizing that behavior may adjust more quickly than deeply-rooted identity.¹⁰ People may be free to adjust their own identity, but the spirit of the concept implies that they do not do this as easily as they can change a pair of socks. To capture this idea we introduce ‘fast’ and ‘slow’ dynamics [see Bergstrom and Lachman 2003]. Specifically, changes in the identity composition of the population (x) are taken to occur slowly relative to changes in cooperative behaviors (z_D and z_c) and the resource stock (S). For example, Bandiera *et al.* [2006] demonstrate that individuals quickly learn how to adapt their behavior (and to adopt a cooperative norm) if it is in their interest. Internalizing such norms, however, likely takes longer.

As above, we start with a general model to outline the main mechanisms, and then switch to a specified model that is analytically tractable. Following Sethi and Somanathan [1996] and OEVG [2007], we use replicator dynamics to describe the quick adjustment of behavior (effort choice), and we also use this approach to model the gradual evolution of identity shares in the population. Specifically, a strategy increases in use when it generates more utility than the average level of utility in society:

¹⁰ Since people are free to eventually update their identity, heterogeneity in terms of a mix of identities is distinct from heterogeneity in terms of variables beyond the control of agents, such as gender, wealth (to some extent), caste, or race. For analyses of the impact of heterogeneity on cooperation, refer to Adhikari and Lovett [2006] and Benin and Pender [2006].

$$(3) \quad \frac{dx}{d\tau} = \alpha_x x (\bar{U}_C(x, z, S) - [x\bar{U}_C(x, z, S) + (1-x)\bar{U}_D(x, z, S)]) \\ = \alpha_x x (1-x) (\bar{U}_C(x, z, S) - \bar{U}_D(x, z, S))$$

$$(4a) \quad \dot{z}_C = \alpha_{zc} z_C (U_{CL}(x, z, S) - \bar{U}_C(x, z, S)) = \\ \alpha_{zc} z_C (1-z_C) (U_{CL}(x, z, S) - U_{CH}(x, z, S))$$

$$(4b) \quad \dot{z}_D = \alpha_{zD} z_D (U_{DL}(x, z, S) - \bar{U}_D(x, z, S)) \\ = \alpha_{zD} z_D (1-z_D) (U_{DL}(x, z, S) - U_{DH}(x, z, S))$$

The ‘dot’ notation in equations (4a,b) means d/dt , where dt refers to a time interval that differs from $d\tau$ (as in (3)). Specifically, we assume $\tau = \varepsilon t$, where $\varepsilon \in (0,1)$ is a fixed parameter, so that behaviors change on a faster time scale than does identity. We return to this concept momentarily. \bar{U}_I refers to the average payoff for agents with identity I .

The dynamic system also includes resource stock dynamics. Assume the resource stock adjusts as a “fast variable” and that growth is quadratic:

$$(5a) \quad \dot{S} = rS(B - S) - [xz_C + (1-x)z_D]E_L S - [x(1-z_C) + (1-x)(1-z_D)]E_H S,$$

where B is the carrying capacity and r is a growth parameter. We earlier defined low effort level E_L as the level that yields greater equilibrium profits than high effort level E_H . It is therefore straightforward to verify that the equilibrium resource stock when everyone chooses low effort, denoted $\bar{S} = B - E_L / r$, must lie on the upward sloping part of the growth function, i.e., $\bar{S} < B/2$ (implying biological overexploitation) and that $E_L > rB/2$.¹¹

The dynamic system consists of four state variables, which can be solved numerically but which would be difficult to analyze using standard phase-plane techniques. We simplify the problem in two ways to make the analysis more tractable. First, note that the condition $U_{DL}(x, z, S) < U_{DH}(x, z, S)$ holds due to our specification in Table 1 (defectors prefer high effort in the absence of regulation). Hence, $z_D = 0$ becomes a globally stable strategy, and equation (4b) may be ignored for now. In this case, $z = z_C$, and (5a) becomes:

¹¹ Having larger equilibrium profits under low effort requires (i) $E_L \bar{S} = r(B - \bar{S})\bar{S} > r(B - \underline{S})\underline{S}$ and (ii) $\bar{S} > \underline{S}$, where $\underline{S} = B - E_H / r$. Using (ii), set $\underline{S} = \bar{S} - \varsigma$, where $\varsigma > 0$ is a parameter, and substitute this relation in for \underline{S} in (i). Simplify the resulting expression so that \bar{S} is by itself, and then take the limit as $\varsigma \rightarrow 0$ to yield $\bar{S} < B/2$. Then solve the equilibrium growth condition, when everyone cooperates, to yield $E_L = r(B - \bar{S}) > r(B - B/2) = rB/2$.

$$(5b) \quad \dot{S} = rS(B - S) - xz_c E_L S - [1 - xz_c] E_H S.$$

Our second simplification stems from the application of Fenichel's Invariant Manifold Theorem [Fenichel, 1979], as our problem meets all three conditions required for its application.¹² Fenichel's theorem allows us to approximate the solution to the fast-slow system by focusing on the simpler problem where $\varepsilon \rightarrow 0$, as the dynamical properties on the slow-time manifold where $\varepsilon \rightarrow 0$ are the same as those on the slow-time manifold where $\varepsilon > 0$. As $\varepsilon \rightarrow 0$, changes in behaviors occur where identity essentially is held fixed.

We focus on the slow-time manifold with $\varepsilon \rightarrow 0$, thereby treating the time frames as infinitely separated. Though we assume that $\varepsilon > 0$ holds, Fenichel's theorem ensures our qualitative results for the identity dynamics are unaffected by this simplification.¹³ Conversely, changes in identity result in seemingly instantaneous behavioral adjustments, from the perspective of time scale τ , as $\varepsilon \rightarrow 0$. This approximation of the true dynamic model has been applied in ecology [Ludwig et al., 1978] and economics [Grimsrud and Huffaker, 2006].

4. ANALYZING THE DYNAMIC MODEL

We proceed by analyzing the dynamics of the fast state variables S and z_c separately from the slow identity dynamics. That is, hold x constant and analyze the system (4a) and (5b). We first derive the $dS/dt=0$ isocline:

$$(6) \quad z_c = \frac{r(B - S) - E_H}{x(E_L - E_H)}.$$

The denominator of (6) is negative, so the $dS/dt=0$ isocline is a positively sloped line in (S, z_c) space. The isocline is depicted in Figure 1, from which we

¹² Define the vector $\mathbf{Y} = [S, z_c]$ and redefine the fast dynamic system by $\dot{\mathbf{Y}} = \mathbf{Q}(\mathbf{Y}, x)$ and $\dot{x} = \varepsilon q(\mathbf{Y}, x)$. The required conditions are as follows: (i) \mathbf{Q} and q are continuous, (ii) a function \mathbf{h} exists such that $\mathbf{Y} = \mathbf{h}(x)$ solves $\dot{\mathbf{Y}} = 0$, and (iii) the real parts of the eigenvalues of $\mathbf{Q}_Y(\mathbf{h}(x), x)$ are all nonzero for all relevant values of x [Fenichel, 1979; Buzzi et al. 2005]. Conditions (i) and (ii) are satisfied by construction, and condition (iii) has been verified numerically.

¹³ The slow-time dynamics with $\varepsilon > 0$ are calculated as asymptotic expansions of the solution on the slow-time manifold with $\varepsilon \rightarrow 0$ [Fenichel, 1979; Kaper and Kaper, 2002]. We limit the slow manifold approximations inserted into slow equation (9) to only the first order terms (the steady state levels), and ignore higher order terms where $\varepsilon > 0$. Kokotovic [1984] analyzes a problem in this fashion and finds that the error is quite small even when ε is significantly larger than zero.

see $z_c \in [0,1]$ for $\underline{S} \leq S \leq S_x$, where $\underline{S} = B - E_H / r$ is the equilibrium stock when no one cooperates and $S_x = (1-x)\underline{S} + x\bar{S} = B - [(1-x)E_H + xE_L] / r$ is the equilibrium stock for the current value of x when all conservationists cooperate. The horizontal phase arrows point towards the $dS/dt=0$ isocline (or, increasing z_c implies effort falls and the stock increases).

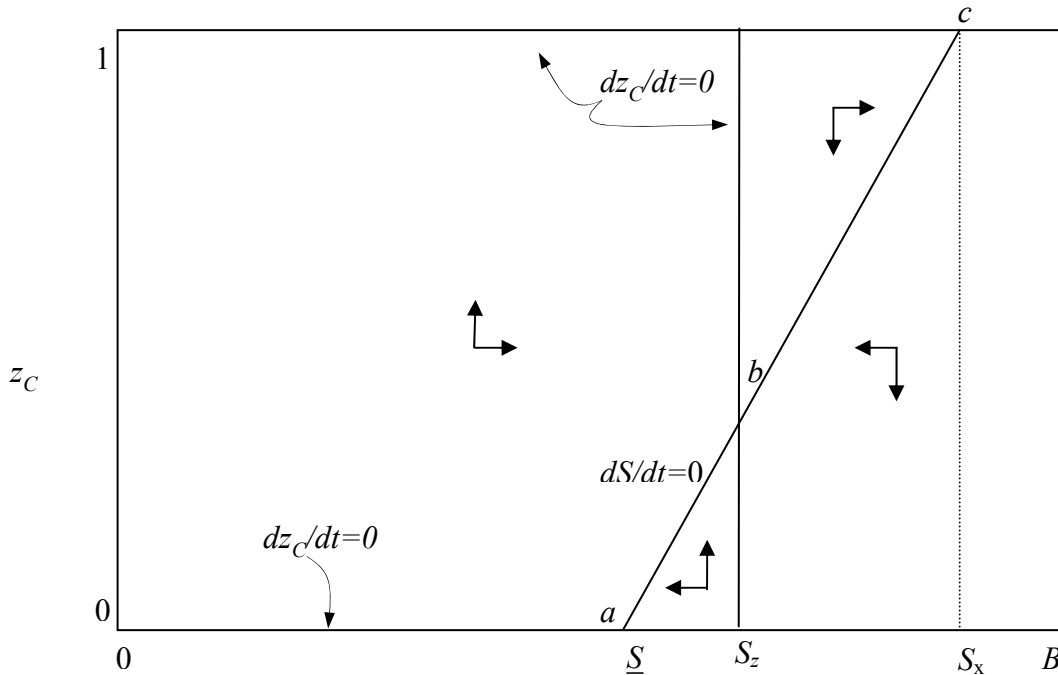


Figure 1a: No outrage costs (no strategic complementarities)

Now consider the dynamics of cooperative behavior. Setting (4a) equal to zero, we obtain two corner equilibria, $z_c = 1$ and $z_c = 0$, and a curve of interior equilibria implicitly defined by:

$$(7) \quad U_{CL}(x, z, S) = U_{CH}(x, z, S).$$

An interior equilibrium occurs when $z = z_c x$ takes on such a value that individuals who have a cooperative identity are indifferent between actually cooperating or not. differentiating (7) gives the slope of the interior isocline:

$$(8) \quad \frac{dz_c}{dS} = \frac{\partial[U_{CH}(x, z, S) - U_{CL}(x, z, S)] / \partial S}{x \{ \partial[U_{CL}(x, z, S) - U_{CH}(x, z, S)] / \partial z \}}.$$

If the marginal value of S is larger upon applying more effort (as assumed in the specified model), the numerator has a positive sign. The denominator should be non-negative by the strategic complementarity property. Hence, the isocline is vertical in the absence of strategic complementarities, and it is positively-sloped in the case of strategic complementarities. The vertical phase arrows point south to the right of the isocline: from (4a), an increase in S , so as to move below the isocline, results in $\dot{z}_c < 0$.

Two phase planes are presented in Figure 1. Figure 1a depicts “no outrage costs” and hence no strategic complementarities. In Figure 1b, outrage costs are of sufficient magnitude to create strong strategic complementarities. In each case, we have assumed the two isoclines cross (cases where the isoclines do not cross are explored below in our discussion of replicator dynamics for x). Thus we get three equilibria: a , b , and c .¹⁴ The stability of these equilibria (in fast time) depends on the magnitude of the strategic complementarities.

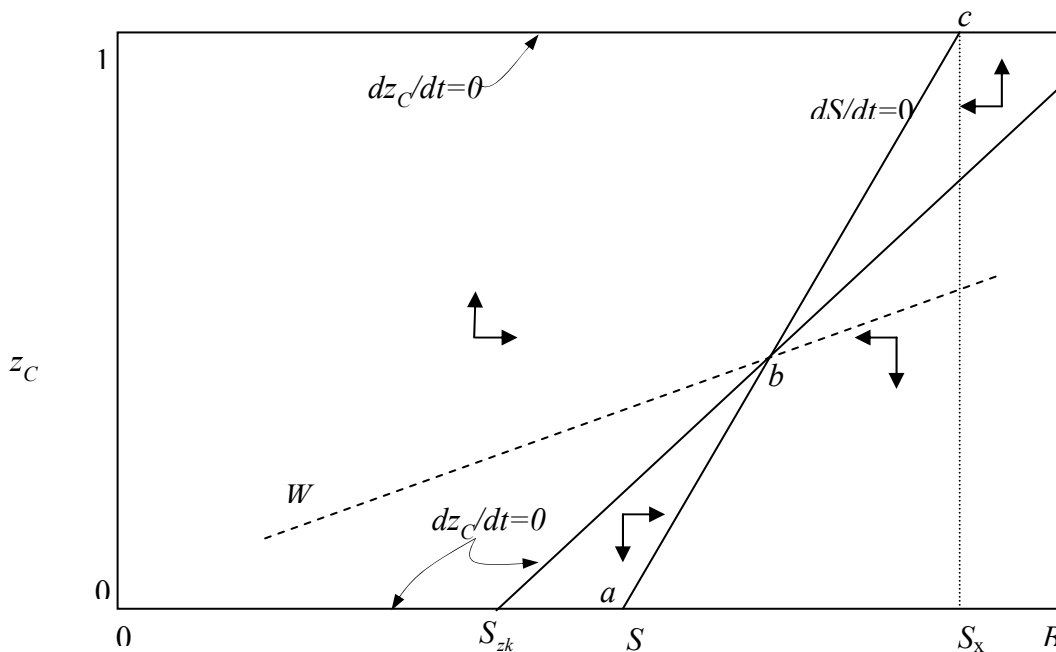


Figure 1b – Outrage costs resulting in strong strategic complementarities

In Figure 1a, equilibria a and c are conditionally stable: they may only be reached for initial values of z_c that are either zero or unity. In contrast, analyzing the eigenvalues of the system learns that equilibrium b is locally stable, and is

¹⁴ An infinite number of equilibria also exist along the vertical axis with $S=0$, but we do not consider this degenerative case as it is unstable provided $\underline{S} > 0$.

either an improper node or a focus point (details available on request). The dynamics are qualitatively the same when strategic complementarities are introduced, provided these complementarities are sufficiently weak that the $dz_c/dt=0$ isocline still cuts the $dS/dt=0$ isocline from below.

In Figure 1b, the strategic complementarities are sufficiently strong causing an upward sloping $dz_c/dt=0$ isocline that cuts the $dS/dt=0$ isocline from above. The stability properties are different in this case, as equilibria a and c are stable while b is a saddle point. The separatrix associated with the saddle is denoted by W . For initial points above W , the system moves to c . For initial points below W , the system moves to a . From this perspective, long-term cooperation is more likely to be sustained if the initial number of people choosing low effort levels is sufficiently high.

However, and unlike OEVG's [2007] model, we find the long run stability of the fast-time equilibria may be compromised as the share of identities, x , in the community slowly evolves according to (3). The evolution of identities depends on which behavioral equilibrium is attained, a , b or c . To clarify this issue, we modify equation of motion (3):

$$(9) \quad \frac{dx}{d\tau} = \alpha_x x(1-x) (\bar{U}_C(x, z^*(x), S^*(x)) - \bar{U}_D(x, z^*(x), S^*(x))),$$

where the relations $z^*(x) = xz_c^*(x)$, and $S^*(x)$ indicate the steady state levels emerging from the fast dynamics, conditional on x . The ultimate equilibrium outcome depends on whether the fast system settles at equilibrium a , b , or c , and possibly also on the initial value of x . A full analysis, however, requires us to return to the specified model. We deal with the cases of no strategic complementarities and sufficiently strong strategic complementarities separately.

4.1 Common pool management with no strategic complementarities

We begin with the case of no outrage costs ($k=g=0$), and therefore no strategic complementarities. The $dz_c/dt=0$ isocline is the vertical line:

$$(10) \quad S = S_z = \frac{w_s^c}{E_H - E_L} > 0.$$

Initially we assume $S_z \in (\underline{S}, S_x)$, so that the dynamics are qualitatively similar to those represented in Figure 1a.¹⁵ Specifically, steady states a and c are conditionally stable, while the interior steady state b is a locally stable, improper node (see appendix). We focus on the case where the system initially moves to equilibrium b , as the other cases are more degenerate and cannot be attained in finite time unless the system begins at the boundary starting values $z_c=0$ or $z_c=1$ (see equations (4a) and (4b)).¹⁶

Equilibrium b in Figure 1a occurs at the point defined by $S^* = S_z$ and $z_c^* = [r(B - S_z) - E_H] / [x(E_L - E_H)]$. This point will slowly change over time because z_c^* depends on x , which evolves at a comparatively slow pace. The evolution of the share of identities are described by plugging the fast-system equilibrium b into equation (9):

$$(11) \quad \frac{dx}{d\tau} = \alpha_x x(1-x) [(w_o^C + w_o^D)x - (w_o^D + w_s^C)].$$

Three equilibria are $x^* = 0$, $x^* = 1$, and $x^* = \hat{x} = [w_o^D + w_s^C] / [w_o^D + w_o^C] > 0$. Note that $\hat{x} < 1$ only occurs when $w_s^C < w_o^C$: it is “less expensive” to act against one’s identity than to interact with individuals with a different identity. If this condition is not satisfied, then $\hat{x} > 1$ and $x^* = 0$ is globally stable for all initial values of $x < 1$. For interior values of \hat{x} the equilibrium \hat{x} must be unstable, while $x^* = 0$ and $x^* = 1$ are each locally stable. Equilibrium $x^* = 0$ will be pursued for initial values of x smaller than \hat{x} , and $x^* = 1$ will be pursued otherwise.

Hence, $\hat{x} \in (0,1)$ has two scenarios ($x \rightarrow 0$ or $x \rightarrow 1$), with multiple forces at play. In addition to the conventional extra harvest benefits for defectors (pulling the system towards $x^*=0$), identity-based benefits and costs might pull the system in the opposite direction. Recall that individuals derive benefits from interacting with their own type. If the number of conservationists is sufficiently large, the fraternal complementarities associated with being a conservationist might dominate the harvest benefits from being a defector, even if the latter yields greater harvest benefits. The initial population share of the two identities, *vis-à-vis* threshold value \hat{x} , thus matters for long run outcomes. Note that these

¹⁵ The case where $S_z > S_x$, comes up in the next section (Figure 2a), and so we delay our discussion of that case for now.

¹⁶ Moreover, if the system happens to start with $z_c=0$ or $z_c=1$, then not even changes in x will move the system away from these equilibrium values; see equation (4a), and also note that changes in x can shift the equilibrium point c horizontally, as we describe below in relation to Figure 2a.

fraternal complementarities generate a self-propelling force. As the share of conservationists in the population increases, welfare of all conservationists is further raised due to enhanced identity benefits, and eventually the population will “snowball” towards a homogenous identity of conservationists. The same applies for increases in the share of defectors.

In the scenario where the share of conservationists in the population declines ($x \rightarrow 0$), defectors acting in accordance with their type are better off than conservationists (who might either act against their type, choosing high effort levels and incurring the associated identity costs, or foregoing the immediate benefits from choosing high effort). As x takes on smaller values, the fast system is affected. The $dS/dt=0$ isocline pivots counter-clockwise on the horizontal intercept $S = \underline{S}$ (see equation 6) so that z_c increases to restore the fast equilibrium: $dS/dt=0$ at $S=S_z$. As the share of conservationists in the population shrinks, holding S constant, conservationist welfare falls regardless of effort choice but remaining conservationists have no immediate incentive to switch effort strategies.

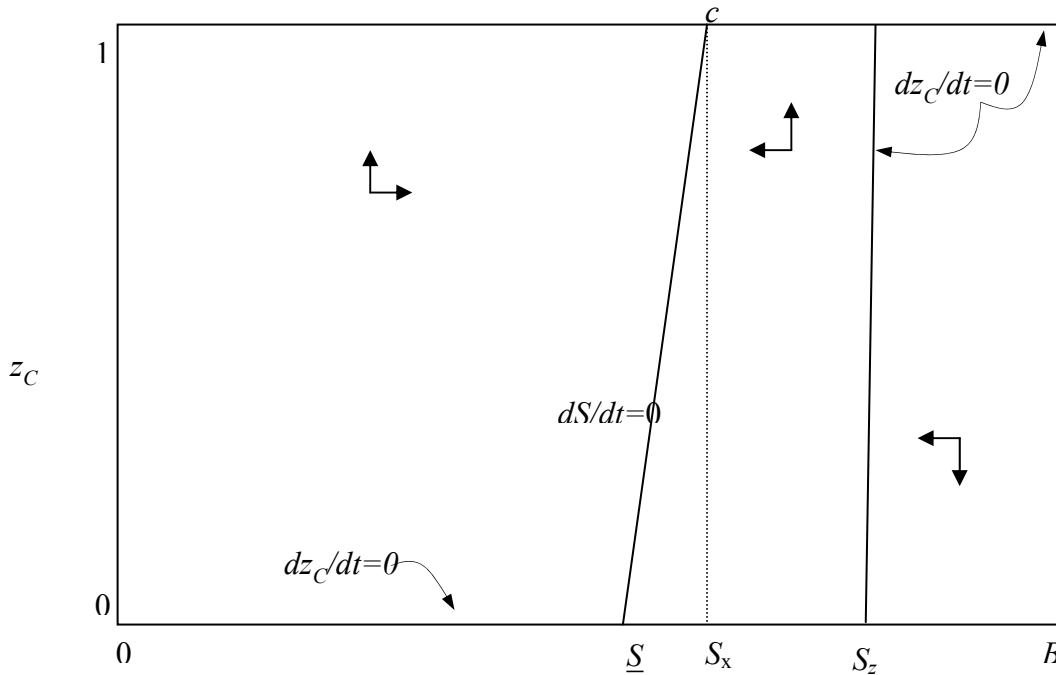


Figure 2a – Isoclines do not intersect (no outrage costs)

However, S starts to fall because everyone who switches to become a defector will now use high effort (note that prior to switching some of the conservationists chose low effort). Once S goes down, the relative gain from

defecting falls so that a larger fraction of the remaining conservationists is inclined to cooperate. At first, the result is that the equilibrium value of S remains constant at S_z , while z_c steadily increases: an increase in defectors leads to greater cooperation among conservationists. But eventually, as x continues to decrease, the isoclines separate (i.e. the interior equilibrium disappears) and the dynamics become as in Figure 2a. All remaining conservationists choose low effort as $z_c=1$ becomes a stable node. But eventually no conservationists are left. The long run equilibrium (when $dx/d\tau = dz/dt = dS/dt = 0$) in this scenario is $S^* = \underline{S}$ and $x^* = 0$.

The scenario $x \rightarrow 1$ represents the opposite case where conservationists are better off than defectors, even in the case where they act against their type and suffer an identity cost. So many conservationists are in the population that they enjoy sufficiently large identity benefits from interacting with their own kind. Initially, when S is constant, conservationists have no incentive to switch effort strategies. However, S starts to rise because not everyone who switches from defector to conservationist uses high effort (as they did prior to the switch). Once S rises, the relative gain from defecting increases and a larger fraction of the conservationists will choose high effort. The result is that the equilibrium value of S remains constant at S_z , while z_c steadily decreases to some positive, equilibrium value. In other words, a decrease in the share of defectors leads to less cooperation among conservationists. The long-term equilibrium of the system is $x^* = 1$, $S^* = S_z$, and $z_c^* = [r(B - S_z) - E_H] / [(E_L - E_H)]$. Society in this case consists only of conservationists, although some members choose low effort and others choose high effort levels. Individuals are indifferent between these choices and thus share identical welfare levels. The high-effort types have larger harvest revenues, but incur offsetting identity costs from acting against their type.

Result 1: *In the absence of outrage costs, conservationists respecting the moral norm of restraint are more (less) likely to cooperate as the number of defectors is increasing (decreasing). The full equilibrium is characterized by either (i) a homogenous population of defectors selecting high effort, resulting in maximum stock depletion, or (ii) a population of conservationists displaying heterogeneous behavior (some choosing high effort, others choosing low effort), resulting in intermediate stock depletion.*

The result of homogeneous identities is different from OEVG. In their model, complementarities act on behavior and allow for a mix of behaviors. The

analogous complementarities in our model act on identity, yet homogeneous identities emerge. However, like OEVG, we still find heterogeneous behaviors are possible in the long run equilibrium.

4.2 Common pool management with strong strategic complementarities

Now consider the scenario where outrage costs are sufficiently large to result in dynamics as in Figure 1b. In addition to the ‘multiple forces’ discussed above, an additional force – strategic complementarities – now pulls the system in a certain direction. The payoffs in Table 1 imply we must rewrite the dynamic equation of the share of cooperators, equation (4a), as:

$$(12) \quad \frac{dz_c}{dt} = \alpha_{zc} z_c (1 - z_c) [w_s^c - (E_H - E_L)S - (1 - xz_c)k],$$

so that the relevant isocline now reads as:

$$(13) \quad z_c = \frac{(E_H - E_L)S - w_s^c + k}{xk}.$$

This isocline is positively-sloped, $\partial z_c / \partial S > 0$, and has a vertical intercept at $S_{zk} = (w_s^c - k) / (E_H - E_L)$. Unlike the earlier specification without outrage costs, the dynamics of behavior are now a function of identity shares in the population. As x increases, the $dz_c/dt = 0$ isocline rotates clockwise, becoming “flatter”. Conversely, as $x \rightarrow 0$, the isocline approaches the vertical line $S = S_{zk}$. From Figure 1b is evident that, again, three potential steady states for the fast-dynamics exist: equilibria a , b and c , where a and c are (locally) stable steady states, and b is a saddle point (see appendix). Because no planner can “place the system on the separatrix”, the probability of arriving at the saddle is essentially zero and may be ignored. Hence $z_c \rightarrow 0$ or $z_c \rightarrow 1$, depending on whether the starting value (S, z_c) is above or below separatrix W . After the fast dynamics have steered the system to either $z_c=0$ or $z_c=1$, it will stay there regardless of what happens to the share of conservationists, x , which changes over time as spelled out in section 4.1. Whether x increases or decreases depends on its initial value relative to a threshold level that depends on the equilibrium value of z_c , denoted $\hat{x}(z_c)$:

$$\hat{x}(1) = \frac{(E_H - E_L) \left(B - \frac{E_H}{r} \right) + k + w_o^D}{w_o^C + w_o^D + k - (E_H - E_L)^2 / r}, \text{ and}$$

$$\hat{x}(0) = \frac{(E_H - E_L) \left(B - \frac{E_H}{r} \right) + w_o^D}{w_o^C + w_o^D}.$$

An initial condition of $x > \hat{x}(z_c)$ implies $x \rightarrow 1$.¹⁷ Conservationists acting in accordance with their type are better off than defectors, because their level of abundance implies that fraternal complementarities (identity benefits) can compensate for lower harvest revenues. In contrast, $x < \hat{x}(z_c)$ implies $x \rightarrow 0$, or the case where defectors are better off so that this type gradually replaces the conservationists.

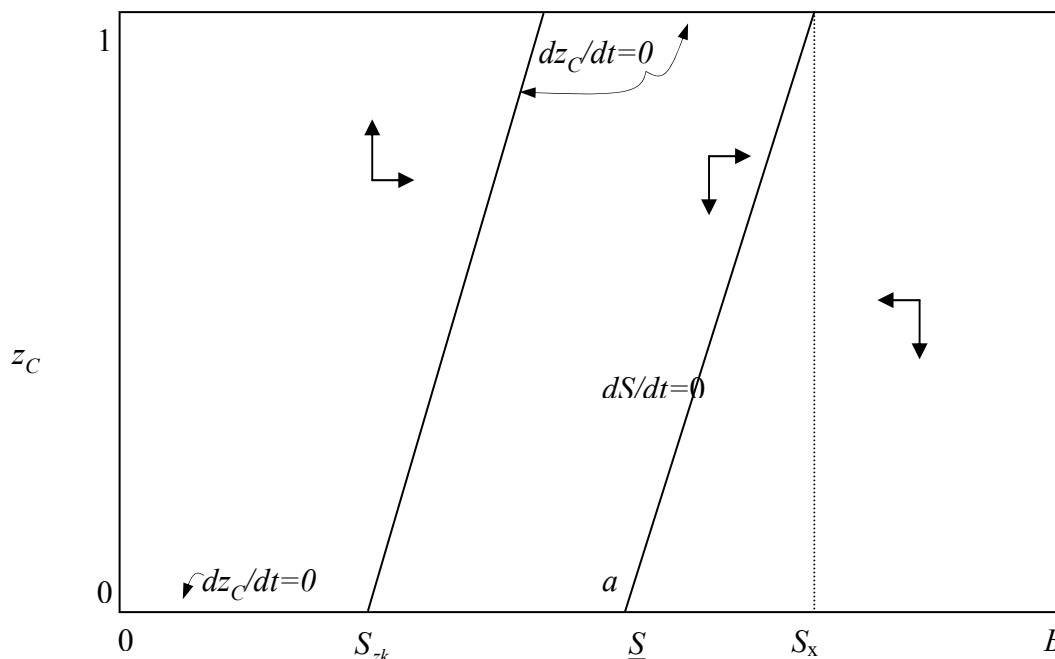


Figure 2b – Isoclines do not intersect (strong strategic complementarities)

As the share of conservationists is updated, the $z_c = 0$ equilibrium will remain locally (or globally) stable as x changes, even if x takes on the value of one. However, the $z_c = 1$ equilibrium could change from being locally stable to conditionally stable as x changes. When $x \rightarrow 0$, both isoclines rotate counterclockwise and eventually separate when x becomes sufficiently small to

¹⁷ Specifically, $\hat{x}(1) = [(E_H - E_L)(B - E_H / r) + k + w_o^D] / [w_o^C + w_o^D + k - (E_H - E_L)^2 / r]$ and $\hat{x}(0) = [(E_H - E_L)(B - E_H / r) + w_o^D] / [w_o^C + w_o^D]$.

weaken the effects of strategic complementarities (since outrage costs depend on both x and k). The dynamics for this case are illustrated in Figure 2b. Here, the equilibrium $z_c = 1$ is conditionally stable while $z_c = 0$ is globally stable for all starting points other than $z_c = 1$. However, one might argue that the value of z_c is immaterial, as $x = 0$ implies nobody cooperates anyway.

Result 2: *In the presence of outrage costs the full equilibrium is characterized by either (i) a homogenous population of defectors disregarding the conservation norm and choosing high effort levels, resulting in maximum stock depletion, or (ii) a homogenous population of conservationists respecting the conservation norm and choosing low effort to conserve the stock.*

The result of homogeneous behaviors is consistent with results in Sethi and Somanathan. In their model and in ours, complementarities act on behavior. But we find complementarities also affect the choice of identities.

5. INTRODUCING REGULATION

We now add regulation to the model in the form of an effort-based fine. The regulator adopts an objective of promoting conservation (i.e., aiming for the “cooperative outcome” by inducing individuals to adopt low harvest effort level E_1), which in our model yields the greatest long run equilibrium benefits. This long run focus is a reasonable alternative to a more complicated model where the regulator also cares about welfare along the path to the long run outcome. Of course a welfare cost is associated with ignoring the transition path, regardless of whether identity-related costs and benefits arise in the model. But an additional long run cost may arise when identity is relevant, which is our focus.

The fine we consider works by punishing deviations from “the cooperative norm.” Since each individual’s identity type is unobservable for the regulator (though the aggregate proportion x may be known), we consider fines that are applied uniformly across identity types. In other words, the regulator cares about behavior, not payoffs or intentions, and anyone choosing high effort rates expects to be fined a penalty F (see Table 1).¹⁸ Regulation has three effects. First, the fine affects relative behavioral payoffs and hence the $dz_c/dt=0$ isocline as conservationists are now more inclined to select low effort levels, *ceteris paribus*.

¹⁸ Qualitatively similar results are obtained when introducing a tax on effort or harvest (in excess of the norm or otherwise).

Second, for a given behavior, the fine affects relative identity payoffs and hence the slow dynamics ($dx/d\tau$). Third, and most importantly, the fine may induce defectors to choose low effort so that $z_D > 0$ is now feasible. A full model, therefore, involves four state variables: z_D , z_C , x and S .

The regulator seeks a fine F to induce both defectors and conservationists to choose low effort, E_L . The minimum penalty that achieves this satisfies (for the more general case of outrage costs):

$$(14) \quad F = \max\{(E_H - E_L)S - w_S^C + (1 - z)k, (E_H - E_L)S + w_S^D + (1 - z)g\}.$$

This critical fine neutralizes the net benefits, in terms of harvest profits as well as the identity costs and benefits, from choosing high effort so that individuals voluntarily switch from high to low effort. How does this fine compare with one that does not account for identity-related payoffs and costs? From (14), the minimum fine that ignored the role of identity would be set according to:

$$(14a) \quad F^* = (E_H - E_L)S.$$

Comparison with (14) indicates that a minimally-set fine that ignores identity would fail to eliminate all high effort choices. Similarly, to compute an optimal fine one needs to know the composition of the population in terms of identities. Regulators can learn about the shares of cooperators and defectors via, for example, economic experiments (e.g., Bouma et al. 2008 use trust games to measure social capital in communities in India).

The critical fine in (14) varies over time as the state of the world changes. The fine is increasing in S , as the opportunity cost of cooperating is increasing in S . In contrast, the fine is decreasing in z : more people cooperating provides greater incentives to others to follow. Once the $z_D = z_C = 1$ equilibrium is achieved, the fine can be dropped (according to (4a) and (4b), the system has reached a permanent outcome).¹⁹ Moreover, $z_D = z_C = 1$ results in $z = 1$, and outrage costs vanish from (14), so the difference between the no-outrage model and the outrage model disappears (i.e. no behavioral externalities). With all of society cooperating on conservation, the stock settles at $S^* = \bar{S} = B - E_L / r$.

¹⁹ This is an artifact of the specification of replicator dynamics, where people switch behavior if they observe that other people are better off. In a more general setting where people can compute an optimal behavioral strategy, the fine F would always be necessary to deter defectors from choosing E_H . In that case the fine can be maintained as a threat (while never actually having to collect on it), and the $z_D = z_C = 1$ equilibrium is locally stable.

At this point, we can restrict the analysis to considering the dynamics of the share of population types x . The fine only indirectly affects these dynamics via its role in attaining the $z_D=z_C=1$ equilibrium. Setting $dx/d\tau=0$, we obtain the interior equilibrium:

$$(15) \quad \tilde{x} = \frac{w_O^D - w_S^D}{w_O^D + w_O^C} < 1.$$

If $\tilde{x} < 0$, then $x = 1$ is globally stable, so that the entire population will eventually consist of conservationists choosing low effort. However, for $0 < \tilde{x} < 1$, then (15) again defines a separatrix: $x > \tilde{x}$ leads to $x=1$ and $x < \tilde{x}$ leads to $x=0$. It can be verified that \tilde{x} is the smallest of the thresholds (i.e., $\tilde{x} < \hat{x}, \bar{x}$). Thus, by affecting harvesting behavior, the fine impacts identity dynamics and increases the likelihood that individuals will become conservationists. However, the fine does not directly impact the threshold \tilde{x} , and so the fine cannot guarantee an all-conservationist outcome ($x=1$). An outcome where all individuals are defectors, choosing low effort E_L , is inevitable for sufficiently low starting values of x .

The long run welfare effects of regulation are necessarily ambiguous. If the fine increases the stock by inducing a behavioral shift from high to low harvest effort, but leaves the identity composition of the community unaffected, then regulation unleashes two opposing forces. First, on the beneficial side, restraint in harvesting effort raises equilibrium harvesting levels (even if harvest effort levels fall). Additional harvests in the steady state represents the conventional component of welfare. But, second, the behavioral switch induced by regulation also leads to identity-based costs whenever $x < \tilde{x}$ – whenever the long-term population consists of defectors forced to act against their type. This cost is derived from individual utility measured in a static sense – representing a “flow” of costs. One instrument cannot generally control two externalities (common property and identity externalities in this case). Assuming that we want to take all preferences – including those of defectors – at face value, then the following observation holds.

Result 3: *An incentive that encourages the appropriate behavior (conservation) by all may not encourage everyone to switch to a conservationist identity and respect conservationist norms. The long-run welfare effects of the regulation are therefore ambiguous in an identity-based model of common pool management. Welfare is reduced when the long-term population consists of defectors who*

act against their type, provided that the associated identity cost w_S^D is sufficiently large to overcome potential harvest gains associated with the cooperative outcome. Even if welfare is increased, it is still lower than it would be in a homogeneous population of conservationists.

Our model thus challenges that unregulated common properties benefit from regulation aimed at stimulating cooperative behavior. The combination of missing property rights and identity effects implies that the standard second-best results apply [e.g., Lipsey and Lancaster, 1956]: addressing one market imperfection (overharvesting due to imperfect property rights) might not improve welfare if the system has other imperfections.

While full cooperation maximizes aggregate long-run profits and promotes stock conservation, it is an open question whether regulators should use a single effort-based tool to strive for this outcome *if they care predominantly about community welfare*. The composition of identity in the population of harvesters matters: if sufficient community members have a conservationist identity, striving for cooperation is efficient. However, when defectors dominate the population, harvesting at the high effort level may maximize welfare. If regulation cannot induce identity shifts, it may be efficient to have systems dominated by defectors go unregulated or else to use additional instruments. This goes back to earlier discussions about needing one policy instrument to address each externality (e.g. Tinbergen 1952), which requires addressing both effort and identity choices in the current model.

Two final observations are relevant here. First, we have argued that regulation yields ambiguous welfare effects when the share of defectors in the population is high (i.e. x is low). But exactly in this context regulators may perceive a need to intervene. While the welfare effects are positive in the presence of sufficient conservationists, this is also the case where intervention might appear redundant. In other words, regulation works when it is really not all that necessary, and might backfire when the decentralized outcome appears particularly bad. Second, an omnipotent government may be able to implement a policy that mimics the social first-best – selecting time-varying policies that address both effort and identity choices, so as to steer the system towards the first-best outcome. However, it is unclear whether this is feasible as identities, unlike behaviors, are not observable.

5.1 Extension: context-dependent identity effects

Until now we have ignored the possibility that punishment can also impact on behavior through changing identity-based benefits and costs. For example, evidence suggests that external intervention may crowd out intrinsic (i.e. identity-based) motivations for norm-abiding behavior. Levitt and Dubner [2005] discuss how payments for blood donations *reduced* supply, and how a children daycare center discovered that introducing fines for late pickups resulted in *more* parents picking up their kids late [see also Gneezy and Rustichini, 2000, for empirical evidence, and Jansen and Mendys-Kamphorst, 2004, for a theoretical model]. Vollaard and Koning [2005] found an inverse relation between a pet tax on dogs and the propensity of dog owners to clean up after their dog. Some evidence suggests that the same interplay between regulation and social capital may apply in the commons [Hatcher et al., 2000; Sutinen et al., 1990].

We capture such phenomena by assuming that a penalty F may relieve conservationists from the guilt they feel when violating the norm: $w_S^C(F)$ with $\partial w_S^C(F)/\partial F < 0$. The fine then ‘crowds out’ the intrinsic motivation to follow (moral) norms, and it is an open question whether monetary incentives or intrinsic motivations do a better job in facilitating cooperation. Alternatively, punishment may mitigate cooperators’ feelings of outrage when meeting a defector. Depending on how we modify the payoff structure, intervention can provide a disincentive to follow the norm. Again, the welfare effects would be ambiguous.

We can easily formalize this a bit, although we leave a full analysis to future work. How does raising an arbitrary fine (smaller than the critical fine derived in (14)), F , impact behavior of conservationists? As an illustration, consider the case of no outrage costs, and assume that prior to intervention the system is in the “heterogeneous equilibrium” discussed in section 4.1 and summarized in Result 1: $x^* = 1$, $S^* = S_{zF}$, and $z_C^* = [r(B - S_{zF}) - E_H]/[(E_L - E_H)]$, where $S_{zF} = (w_S^C + F)/(E_H - E_L)$. The system is in locally stable equilibrium b (see Fig 1a), and the effect of regulation on conservation is determined by movements of the relevant isoclines across the phase plane. Note that, for an arbitrary fine F , the $dz_c/dt = 0$ isocline, as well as the equilibrium resource stock, are described by $S^* = S_{zF}$. The impact of intervention on this isocline is as follows:

$$(16) \quad \left. \frac{dS_{zF}}{dF} \right|_{dz_c/dt=0} = \frac{\left(\frac{\partial w_S^C}{\partial F} \right) + 1}{E_H - E_L}.$$

Hence, for $\partial w_S^C / \partial F < -1$, the isocline shifts to the left, and the resource stock will become smaller. More conservationists act against their type, because the identity-based costs associated with a-typical behavior are smaller as formal rules have crowded out informal rules of restraint. Hence, the effects of regulation are ambiguous for behavior, welfare and stock conservation.

6. CONCLUSIONS, DISCUSSION AND RECOMMENDATIONS

The management of common property resources receives ample academic attention. Economists explore the efficiency and equity considerations of communal management, and the role of common resources as a safety net for the poor has been extensively discussed. Yet, experiences with respect to the management of CPR are diverging, and much remains unknown about the conditions under which CPR are sustainably managed. This paper tries to contribute to a better understanding of CPR management by analyzing the interaction between resources, harvesting effort, and the evolution of social norms.

Building on the identity model of Akerlof and Kranton, we distinguish between beliefs (values) and behavior.²⁰ We analyze the implications of a heterogeneous set of beliefs and values among resource users, and allow individuals to slowly update their beliefs and values in response to economic pressure. The latter feature represents a methodological contribution, and generates a model where people self-select into a set of preferences. We view the share of resource users ‘believing’ in norms of harvesting restraint as a proxy of social capital. Hence our model allows a systematic analysis of the accumulation and erosion of social capital in the commons.

Adjusting one’s deep-seated beliefs is a relatively slow process—slower than adjusting one’s behavior. The potential mismatch between behavior and beliefs sets the stage for various dynamic adjustment processes. We distinguish between conventional ‘harvest payoffs’ and ‘identity-derived payoffs’ following from own behavior and behavior of others. Identity-based payoffs are non-pecuniary costs emerging when acting against one’s ‘internalized deep values’ (or observing others violating such norms). Identity-based benefits also emerge when individuals interact with others with whom they share these deep values, so that total payoffs of behavioral strategies depend on both own and others behavior, as

²⁰ Distinguishing between identity and behavior is important as highlighted by a littering study by Cialdini et al. (1993). While littering behavior is influenced more by social attitudes (norms) towards littering rather than the individual’s perception of what is typically done in a specific setting, Cialdini et al. document that littering behavior is influenced by the actions of others. Individuals can act against a norm they have internalized.

well as the composition of the population in terms of the distribution of norms. We demonstrate that in the absence of formal or informal enforcement, such a system is characterized by multiple equilibria, and identify conditions under which social capital is ‘built up’ or ‘broken down.’ We also demonstrate that regulation aimed at internalizing harvest externalities, in a context where identities matter, generally produces ambiguous welfare effects. Local welfare may benefit or suffer from penalizing overharvesting.

What is the contribution to the literature? First, the paper provides a conceptual contribution to the CPR literature. Experimental evidence (e.g., Andreoni 1988, Fischbacher et al. 2001) and case studies (e.g., Feeny et al. 1996, Baland and Platteau 1996) suggest individuals are motivated by a range of factors, including non-pecuniary rewards and altruism. Many agents respect sharing norms, taboos, and norms encouraging a sustainable livelihood. The model structure in this paper provides a framework to analyze the interaction between norms, behavior and economic outcomes. Second, the model allows studying the welfare implications of interventions when people can adjust both their behavior as well as their beliefs. Third, the identity model generates a wealth of stable equilibria, encompassing outcomes of earlier work as ‘special cases.’ For example, earlier work finds equilibria where everybody respects or rejects norms of restraint, or predicts equilibria with ‘mixed behaviors.’ In contrast, our model outlines conditions under which heterogeneous and homogenous stable equilibria will occur—outcomes with mixed identities and behaviors, or outcomes where all agents are of the same type and act the same, respectively.

Incorporating identity-based costs and benefits to the domain of resource economics sets the stage for a range of interesting analyses. One avenue for future work to improve the current analysis is to incorporate decentralized punishment. Another interesting area for future work is formally deriving welfare maximizing regulations in the presence of identity effects.

APPENDIX

A. Stability of the interior equilibrium, b , in the case of no outrage costs

The two equations of motion are:

$$(A1) \quad \dot{z}_C = \alpha_{zc} z_C (1 - z_C) ([E_L - E_H] S + w_s^c), \text{ and}$$

$$(A2) \quad \dot{S} = rS(B - S) - xz_C E_L S - [1 - xz_C] E_H S.$$

To determine the eigenvalues, we want to linearize these equations around the equilibrium point b , which is

$$(A3) \quad z_C^* = \frac{r(B - S^*) - E_H}{x(E_L - E_H)}, \quad S^* = S_z = \frac{w_s^c}{E_H - E_L}.$$

Using Mathematica 7.0 (Wolfram 2008) to derive the eigenvalues for this system, we obtain:

$$(A4) \quad \frac{1}{2} \left(-rS_z \pm \sqrt{r^2 S_z^2 - 4\alpha_{zc} [E_H - E_L]^2 S_z x (1 - z_C^*)} \right),$$

where the first term in parentheses is negative and the second term in parentheses is either (i) imaginary, or (ii) real but less than rS_z . Case (i) results in a stable focus, while case (ii) results in a stable node.

B. Stability of the interior equilibrium, b , in the case of outrage costs

The two equations of motion are:

$$(B1) \quad \dot{z}_C = \alpha_{zc} z_C (1 - z_C) (w_s^c - [E_H - E_L] S - (1 - x z_C) k), \text{ and}$$

$$(B2) \quad \dot{S} = rS(B - S) - x z_C E_L S - [1 - x z_C] E_H S.$$

To determine the eigenvalues, we want to linearize these equations around the equilibrium point b , which is

$$(B3) \quad z_C^* = \frac{(E_H - E_L) E_H - r(b[E_H - E_L] - w_s^c + k)}{x[(E_H - E_L)^2 - rk]},$$

$$S^* = \frac{(E_H - E_L) w_s^c + k(E_L - br)}{(E_H - E_L)^2 - rk}.$$

Using Mathematica 7.0 (Wolfram 2008) to derive the eigenvalues for this system, we obtain:

$$(B4) \quad \frac{1}{2} \left(A \pm \sqrt{A^2 + \alpha_{zc} z_C (1 - z_C) 4(rk - [E_H - E_L]^2 x) S^*} \right),$$

where $A = -rS^* + \alpha_{zc}kz_c^*(1 - z_c^*)$. The discriminant in expression (B4) is positive and greater than A^2 for the phase plane illustrated in Figure 1b. In particular, the second term in the discriminant is positive as $x < 1$ and as $rk > [E_H - E_L]^2$ when the $dS/dt = 0$ isocline cuts the $dz_c/dt = 0$ isocline from below, as in Figure 1b. Hence, the square root term in (B4) is greater than A , resulting in one positive and one negative eigenvalue. This means the equilibrium point b must be a saddle.

REFERENCES

- Adhikari, B. and J. Lovett, 2006. Institutions and Collective Action: Does Heterogeneity Matter in Community-Based Resource Management, *Journal of Development Studies* 3, 426-445.
- Ahuja, V., 1998. Land Degradation, Agricultural Productivity and Common Property: Evidence from Cote d'Ivoire, *Environment and Development Economics* 3, 7-34.
- Akerlof, G.A. and R.E. Kranton, 2000. Economics and Identity, *Quarterly Journal of Economics* 115, 715-753.
- Akerlof, G.A. and R.E. Kranton, 2002. Identity and Schooling: Some Lessons for the Economics of Education, *Journal of Economic Literature* 40, 1167-1201.
- Akerlof, G.A. and R.E. Kranton, 2005. Identity and the Economics of Organizations, *Journal of Economic Perspectives* 19, 9-32.
- Baland, J. and J.P. Platteau, 1996. *Halting Degradation of Natural Resources: Is there a Role for Rural Communities?* Oxford: Clarendon Press.
- Baland, J.M. and P. Francois, 2005. Commons as Insurance and the Welfare Impact of Privatization, *Journal of Public Economics* 89, 211-231.
- Bandiera, O., I. Barankay, and I. Rasul, 2006. The Evolution of Cooperative Norms: Evidence from a Natural Field Experiment, *Advances in Economic Analysis & Policy* 6:
- Benin, S. and J. Pender, 2006. Collective Action in Community Management of Grazing Lands: The Case of the Highlands of Northern Ethiopia, *Environment and Development Economics* 11, 127-150.
- Bergstrom, C.T. and M. Lachman, 2003. The Red King Effect: When the Slowest Runner Wins the Coevolutionary Race, *Proceedings of the National Academy of Sciences* 100, 593-598.
- Bison, A. and T. Verdier, 2001. The Economics of Cultural Transmission and the Dynamics of Preferences. *Journal of Economic Theory* 97, 298-319.
- Brekke, K.A., S. Kverndokk and K. Nyborg, 2003. An Economic Model of Moral Motivation, *Journal of Public Economics* 87, 1967-1983

- Brito, D., M.D. Intriligator, E. Sheshinski, 1997. Privatization and the Distribution of Income in the Commons, *Journal of Public Economics* 64, 181-205.
- Becker, G., 1976. *The Economic Approach to Human Behavior*, Chicago: University of Chicago Press.
- Bouma, J., E. Bulte and D. van Soest, 2008. Trust, Trustworthiness and Cooperation: Social Capital and Community Resource Management, *Journal of Environmental Economics and Management* 56, 155-166.
- Buzzi, Da Silva and Teixeira, 2005. Single Perturbation Problems for Time Reversible Systems. *Proceedings of the American Mathematical Society* 133, 3323-3331.
- Cialdini, R., R. Reno and C. Kallgren, 1993. The Transituational Influence of Social Norms, *Journal of Personality and Social Psychology* 64, 104-112
- Dekel, E., J.C. Ely and O. Yilankaya, 2007. Evolution of Preferences, *Review of Economic Studies* 74, 685-704.
- De Meza, D. and J.R. Gould, 1987. Free Access versus Private Property in a Resource: Income Distributions Compared, *Journal of Political Economy* 95, 1317-1325.
- Dowell, R.S., R.S. Goldfarb and W.B. Griffith, 1998. Economic Man as a Moral Individual, *Economic Inquiry* 36, 645-653.
- Dutta, P., 1995. Collusion, Discounting and Dynamic Games. *Journal of Economic Theory* 66, 289-306.
- Fafchamps, M., 2004. *Market Institutions in Sub-Saharan Africa: Theory and Evidence*, Cambridge: MIT Press.
- Feeny, D., S. Hanna and A.F. McEvoy, 1996. Questioning the Assumptions of the "Tragedy of the Commons" Model of Fisheries, *Land Economics* 72, 187-205.
- Fehr, E. and S. Gächter, 2000. Fairness and Retaliation: The Economics of Reciprocity, *Journal of Economic Perspectives* 14, 159-181.
- Fenichel, N., 1979. Geometric Singular Perturbation Theory for Ordinary Differential Equations, *Journal of Differential Equations* 31, 53-98.
- Fischbacher, U., S. Gächter and E. Fehr, 2001. Are People Conditionally Cooperative? Evidence from a Public Goods Experiment, *Economics Letters* 71, 397-404.
- Francois, P. and J. Zabojsnik, 2005. Trust, Social Capital and Economic Development, *Journal of the European Economic Association* 3, 51-94.
- Glaeser, E.L., D. Laibson and B. Sacerdote, 2002. An Economic Approach to Social Capital, *The Economic Journal* 112, 437-458.
- Glaeser, E.L., B. Sacerdote and J.A. Scheinkman, 2003. The Social Multiplier, *Journal of the European Economic Association* 1, 345-353.

- Grimsrud, K.M. and R. Huffaker, 2006. Solving Multidimensional Bioeconomic Problems with Singular-Perturbation reduction Models: Application to Managing Pest Resistance to Pesticidal Crops. *Journal of Environmental Economics and Management* 51, 336-353.
- Gordon, H.S., 1954. The Economic Theory of a Common Property Resource: The Fishery, *Journal of Political Economy* 62, 124-142.
- Gneezy, U. and A. Rustichini, 2000. A Fine is a Price, *Journal of Legal Studies* 29, 1-18.
- Guth, W. and M. Yaari, 1992. Explaining Reciprocal Behavior in Simple Strategic Games: An Evolutionary Approach. In: U. Witt (ed.), *Explaining Process and Change: Approaches to Evolutionary Economics*. Ann Arbor: University of Michigan Press.
- Hannesson, R., 1997. Fishing as a Supergame, *Journal of Environmental Economics and Management* 32, 309-322.
- Haslam, S.A., 2001. *Psychology in Organizations: the Social Identity Approach*. Thousand Oaks: Sage Publications.
- Hatcher, A., S. Jaffry, O. Thébaud and E. Bennett, 2000. Normative and Social Influences affecting Compliance with Fishery Regulations, *Land Economics* 76, 448-461.
- Henrich, J., R. Boyd, S. Bowles, C. Camerer, E. Fehr, H. Gintis, and R. McElreath, 2001. In Search of Homo Economicus: Behavioral Experiments in 15 Small-Scale Societies, *American Economic Review (Papers & Proceedings)* 91, 73-78.
- Henrich, J., 2004. Cultural Group Selection, Coevolutionary Processes and Large Scale Cooperation, *Journal of Economic Behavior & Organization* 53, 3-35.
- Horst, U. and J. Scheinkman, 2003. Equilibria in Systems of Social Interactions. *Princeton Economic Theory Working Paper Series*, Princeton: New Jersey, U.S.A.
- Janssen, M.C.W. and E. Mendys-Kamphorst, 2004. The Price of a Price: On the Crowding Out and In of Social Norms, *Journal of Economic Behavior & Organization* 55, 377-395.
- Kandori, M., 1992. Social Norms and Community Enforcement, *Review of Economic Studies* 59, 63-80.
- Kaper, H. and T. Kaper, 2002. Asymptotic Analysis of Two Reduction Methods for Systems of Chemical Reactions, *Physica D* 165, 66-93.
- Kokotovic, P.V., 1984. Applications of Singular Perturbation Techniques to Control Problems, *SIAM Review* 26, 501-550.
- Kranton, R., 1996. The Formation of Cooperative Relationships, *Journal of Law, Economics and Organizations* 12, 214-233.
- Krugman, P., 1991. History versus Expectations, *Quarterly Journal of Economics* 106, 651-667.

- Kuran, T. and W. Sandholm, 2008. Cultural Integration and its Discontents. *Review of Economic Studies* 75, 201-228.
- Levitt, S.D. and S.J. Dubner, 2005. *Freakonomics*, New York: Harper-Collins.
- Lindbeck, A., 1997. Incentives and Social Norms in Household Behavior, *American Economic Review* 87, 370-377.
- Lipsey, R.G., and K. Lancaster, 1956. The General Theory of Second Best. *Review of Economic Studies* 24, 11-32.
- Lopez, R., 2001. The Tragedy of the Commons in Cote d'Ivoire Agriculture: Empirical Evidence and Implications for Evaluating Trade Policies, *World Bank Economic Review* 12, 105-131.
- Ludwig, D., D. Jones and C. Holling, 1978. Qualitative Analysis of Insect Outbreak Systems: The Spruce Budworm and Forest, *Journal of Animal Ecology* 47, 315-332.
- Mailath, G. and L. Samuelson, 2006. *Repeated Games and Reputation*. Oxford: Oxford University Press.
- McCarthy, N., E. Sadoulet and A. de Janvry, 2001. Common Pool Resource Appropriation under Costly Cooperation, *Journal of Environmental Economics and Management* 42, 297-309.
- Nowak, M., 2006. Five Rules for the Evolution of Cooperation, *Nature* 314, 1560-1563
- Nyborg, K., R. Howarth and K.A. Brekke, 2006. Green Consumers and Public Policy: On Socially-Contingent Moral Motivation, *Resource and Energy Economics* 28, 351-366
- Osés-Eraso, N. and M. Viladrich-Grau, 2007. On the Sustainability of Common Property Resources, *Journal of Environmental Economics and Management* 53, 393-410.
- Ostrom, E. (1990), *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge: Cambridge University Press.
- Ostrom, E., R. Gardner, and J. Walker, 1992. Covenants With and Without Sword: Self-Governance is Possible, *American Political Science Review* 86, 404-417.
- Ostrom, E., Gardner, R., and Walker, J. (1994), *Rules, Games and Common-Pool Resources*, Ann Arbor: The University of Michigan Press.
- Platteau, J.P., 1994. Behind the Market Stage where Real Societies exist: Part II, The Role of Moral Norms, *Journal of Development Studies* 30, 753-815.
- Polasky, S., T. Narui, G.M. Ellis, C.F. Mason, 2006. Cooperation in the Commons. *Economic Theory* 29, 71-88.
- Rice, S.H., 2004. *Evolutionary Theory: Mathematical and Conceptual Foundations*. Sunderland, MD: Sinauer.
- Sethi, R. and E. Somanathan, 1996. The Evolution of Social Norms in Common Property Resource Use, *American Economic Review* 86, 766-788.

- Sutinen, J.G., A. Rieser and J.R. Gauvin, 1990. Measuring and Explaining Noncompliance in Federally Managed Fisheries, *Ocean Development and International Law* 21, 335-372.
- Taifel, H., 1978. Interindividual Behavior and Intergroup Behavior. In: H. Taifel (ed.): *Differentiation between Social Groups: Studies in the Social Psychology of Intergroup Behavior*. London: Academic Press.
- Tarui, N., 2007. Inequality and Outside Options in Common-Property Resource Use, *Journal of Development Economics* 83, 214-239.
- Tinbergen, J., 1952. *On the Theory of Economic Policy*. Amsterdam: North Holland
- Vendrik, M., 1993. *Collective Habits and Social Norms in Labour Supply: From Micro Motives to Macro Behavior*. Dissertation 93-18. Maastricht: University of Limburg
- Vollaard, B. and P. Koning, 2005. Meer Hondenontlasting door Hondenbelasting ("More Dog Shit because of Dog Taxation"), *Economisch Statistische Berichten*: 296-297.
- Watson, J., 1999. Starting Small and Renegotiation. *Journal of Economic Theory* 85, 52-90.
- Weitzman, M., 1974. Free Access versus Private Property as Alternative Systems for Managing Common Property, *Journal of Economic Theory* 8, 225-234.