

The survival of the conformist: equity-driven ostracism and renewable resource management

Alessandro Tavoni^{1,2}, Maja Schlüter², Simon Levin²

¹ Advanced School of Economics at the University of Venice, Cannaregio 873,
30121 Venice, Italy

² Department of Ecology and Evolutionary Biology, Princeton University,
Princeton, NJ 08544-1003

Alessandro Tavoni, atavoni@princeton.edu
Maja Schlüter, mschlute@princeton.edu
Simon Levin, slevin@princeton.edu

Abstract. This paper examines the role of pro-social behavior as a mechanism for the establishment and maintenance of cooperation in resource use under variable social and environmental conditions. By coupling resource stock dynamics with social dynamics concerning compliance to a social norm prescribing non-excessive resource extraction in a common pool resource (CPR), we show that when reputational considerations matter and a sufficient level of social stigma affects the violators of a norm, sustainable outcomes are achieved. We find large parameter regions where norm-observing and norm-violating types coexist, and analyze to what extent such coexistence depends on the environment.

Keywords: Other-regarding behavior, cooperation, social norm, ostracism, common pool resource, evolutionary game theory, replicator equation, agent-based simulation, coupled socio-resource dynamics

1. Introduction

Local and global ecosystems are under growing pressure worldwide, and beyond any doubt their sustainable management cannot be achieved without the stakeholders' cooperative efforts. History has taught us that the livelihood of our species is inextricably related to our ability to cooperate, in the sense of restraining use of natural resources to sustainable levels, rather than giving in to excessive resource appropriation. However, depending on the characteristics of the system at hand, tensions between individual and collective good may undermine such norm of restraint. CPRs where beneficiaries of the resource have open access to it are a notable case of appropriation externality paving the way for short-sighted resource utilization: all agents would be better off if they collectively restrained extraction, but if the impact of one's action on the resource stock is ignored, it is individually rational not to do it. Thus, maintaining cooperation against the myopic self-interest of a potentially large fraction of individual users unwilling to restrict their behavior for the collective good, and despite growing environmental pressure, is a challenging task that often depends on a multitude of factors, as both successful and unsuccessful environmental management has shown. Nevertheless, field work, controlled experiments involving participants playing stylized games aimed at reflecting the trade-offs inherent in these social dilemmas (e.g., CPR and public good games), as well as casual observation, suggest that human beings are able to overcome the obstacles to cooperation in a variety of settings. Many explanations have been proposed to account for the widely-observed departures from the rational-agent models' predictions of collectively inefficient resource management in the absence of regulatory institutions. Established mechanisms that have been advanced to account for the evolution of cooperation are, following Nowak (2006): kin selection (the inclination of related individuals to engage in cooperative behavior), direct reciprocity (the "I will scratch your back if you scratch mine"

attitude towards reciprocating), indirect reciprocity (I will scratch your back because someone else scratched mine), network reciprocity (spatial structure is assumed to allow for unevenly mixed populations where some individuals interact more frequently than others) and multilevel selection (where the population is divided into groups whose members are allowed to enact different strategies depending on whether they are matched with own-group members or with members of other groups).¹ These mechanisms have been shown to suffice for the evolution of cooperation in Prisoner's Dilemma games whenever the payoffs are such that the benefit-to-cost ratio of the cooperative action exceeds a certain mechanism-specific threshold. While the above mechanisms incorporate some of the empirically observed factors that influence the success of collective action, such as the topology of interactions and group size, we postulate that the link to other important drivers of cooperation needs to be made explicit if one wants to attempt to bridge the gap between the empirical findings on commons management and the theory. In the present paper we aim to analyze a simple model that departs from the full-rationality paradigm placing emphasis on two such drivers: the presence of individuals with other-regarding preferences and the conformist pressure in the direction of norm compliance arising from fear of community disapproval. Laboratory experiments, such as Fehr and Fischbacher (2002) and Maier-Rigaud et al. (2008), respectively suggest that both are relevant, while contributions from social psychology (Cialdini, 1984) and the empirical literature on the commons stress the importance of the second driver. For what concerns the empirical findings, the work of Ostrom (1990 and 2007) has suggested that many CPRs have escaped the trap of the tragedy of the commons by being managed in a self-organizing manner with mechanisms such as rules, norms and graduated sanctions favoring cooperation among users. Such mechanisms may result from the repeated interactions of the resource users: given suitable conditions, for example in terms of knowledge of the ecosystem by the community members, they could develop a sense of what behavior is acceptable with respect to resource use. The enforcement of the proper behavior or sanctioning thereof will take different forms depending on the features of the social and resource system at hand. One increasingly studied class concerns social norms. Following the work of Sethi and Somanathan (1996), much attention has been given to the role of costly punishment of defectors in promoting cooperation; recent contributions aimed at extending their setup have been proposed by Noailly et al. (2007) and Sethi and Somanathan (2006) While both retain the three agent types format, with defectors, cooperators and enforcers bearing the costs of punishment, the former allow for spatial structure in the interactions, and the latter introduce a concern for reciprocity among the agents. Yet, the empirical literature on the commons argues that a variety of sanctioning mechanisms against norm violators are utilized to promote successful management of irrigation systems, fisheries, pastures and forests (Ostrom 1990, Baland and Platteau 1996, ch.8 and 11). Moreover, as suggested by the literature on social capital (Bowles and Gintis 2002, Osés-Eraso and Viladrich-Grau 2007, Iwasa 2009), resource appropriators embedded in a social context can often rely on a wider set of tools than the traditionally considered costly sanctioning

¹ More recently the term of assortment, indicating the "degree of segregation of different types of individuals into different groups" as Pepper (2007) puts it, has gained consensus among scholars for its generality. See van den Bergh and Gowdy (2009) and Fletcher and Doebli (2009) for recent contributions to the group selection debate. It should be noted that many other mechanisms with a certain degree of overlapping characteristics have been employed in various disciplines to highlight the tension between in-group and outsiders; among others, parochialism (Choi and Bowles 2007,) and homophily (Currarini et al. 2009).

of free-riding behavior. When the result of one's actions is observable, be it the resource extraction itself or the outcome of a productive activity which is dependent on the latter, field and experimental evidence suggests that individuals belonging to a community act more cooperatively than when in isolation, as a result of their exposure to social reprobation.² In the present paper we focus on one such mechanism, which we term equity-driven ostracism. The underlying idea is that appropriators' decisions about how much effort to exert in the extraction of a natural resource are based on the prevailing norms that have emerged in the community, in addition to the usual efficiency considerations. As a result of the compliance decision with respect to the norm, those who deviate (the defectors) may be refused resources and support by those who comply. As an example, the ostracism costs considered here could be thought of as originating from destruction of defector's crop by the cooperators, or simply from refusal of help by the community towards a defector in the form of denial of loan of machinery or means of transportation needed to take the harvest to the market. The rationale for this behavior is that, in a community of individuals who share access to a natural resource, those who restrain their extraction level to the socially acceptable level will not show the same level of support they have for fellow cooperators, when it comes to defectors. The inherent tradeoff between unrestricted profit seeking and norm adherence can be visualized as follows:

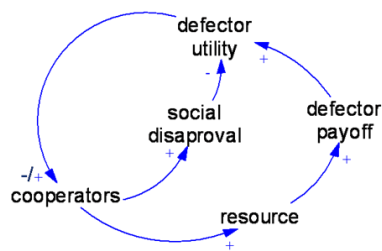


Figure 1: Interactions between the composition of the population, resource abundance and norm enforcement

The above schematization allows us to highlight the ever-changing conditions faced by appropriators choosing between extraction patterns. If the number of cooperators increases, so will the resource stock, favoring the defectors the most (due to their behavior being unrestricted by the norm) and rendering opportunistic behavior more profitable, thus posing unfavorable selective pressure on the cooperators' population. These trends are captured by the outer arrows originating from the cooperators in Figure 1, with favorable (unfavorable) changes marked with a "+" ("−") sign to the right of the arrows. However, as shown by the inner arrows, an increase in the number of cooperators also leads to greater social disapproval towards norm violators, as fewer defectors will face ostracism by a larger community of norm followers denying them the benefits of cooperation.

² According to Gowdy (2008): "Experimental results from behavioral economics, evolutionary game theory and neuroscience have firmly established that human choice is a social, not self-regarding, phenomenon. [...] Human decision-making cannot be accurately predicted without reference to social context". Recent evidence on the importance of the social context in guiding individual behavior is found in Fehr and Fischbacher (2002) and Akerlof (2007).

We implement this simple mechanism in an evolutionary framework in order to allow for departures from optimizing behavior as prescribed by Nash equilibrium. Namely, rather assuming that the resource appropriators instantaneously maximize their material well-being in a repeated-interaction model with discounted future, we let evolution gradually shape the proportion of agents playing a given strategy by favoring the more successful one. The advantage of this bounded rationality approach (imitate the successful behavior with inertia) is that it avoids the downfalls of the multiplicity of equilibria and lack of robustness to noise, while retaining the behavioral tendency to move in the direction of a more profitable strategy.³ The results of the analysis, presented in section 2, suggest that both monomorphic and polymorphic populations emerge: that is we find stable full compliance (all-C) and full defection (all-D) equilibria as in Sethi and Somanathan (1996), but also mixed equilibria where both types coexist. In section 3 we investigate the impact of variation in environmental conditions, by allowing for variability in the rate of resource regeneration as well for a multiplicity of extraction strategies subject to evolutionary pressure. Section 4 provides conclusive remarks. .

2. A model of coupled socio and resource dynamics

We examine the role of other-regarding behavior as a mechanism for the establishment and maintenance of cooperation in resource use under variable social and environmental conditions. This is done by modeling the evolution of compliance to a social norm prescribing conformity to an agreed extraction level, and its coevolution with a CPR stock dynamics. The coupled dynamics allows us to investigate the stability of cooperation in a population of resource users who have symmetrical access to it and are not only concerned about own yield from productive use of the resource, but also about their status with respect to other community members, as acceptance to the community is at stake. Payoff comparisons (e.g. with respect to crop size) lead to ostracism of the norm violators by the cooperating community, which denies defectors the benefits of resource and knowledge support, imposing losses on them that may offset those incurred by cooperators when restricting resource extraction practices to more sustainable ones. That is, individuals face a trade-off: on the one hand they can extract resource at individually optimal (but socially detrimental) levels without restricting usage, or on the other hand they can constraint themselves to a socially agreed upon acceptable level. By doing so, their conventional materialistic pay-off is necessarily below that of the non-cooperating agents, because of the above mentioned lower extraction and consequently reduced yield (which is increasing in the extraction effort). However, violators of the social norm, insofar as they are recognized as such, are penalized by being excluded from the help of the cooperating community (e.g. in bringing the yield to the market): such defectors-specific ostracism costs have a variable magnitude that depends on the relative size of the cooperating community, since at low frequencies of cooperative agent types, the defectors will be better off, but at high enough frequencies of cooperators, the former may incur high enough ostracism costs that it will be advantageous to be part of the sustainable community. Lindbeck (1997) suggests this network good property of norms: “a social norm is felt more strongly, the greater the number of individuals who obey it. Thus, the adherence to a social norm is a

³ For a critique of the commonly used approaches in the economic analysis of common property, see Sethi and Somanathan (2006).

choice conditioned on other individuals' adherence to the same norm. The psychological explanation for this type of behavior may be either that disapproval from others is more troubling if expressed by many people than by few or that other people's behavior is assumed to signal information about what is proper or potentially successful behavior.”

Agents are considered as productive units (one can think of an agent as an individual or a family), whose share of the total production (e.g. the size of their crop) is proportional to the share of their appropriation effort with respect to the aggregate effort. Their source of revenue is assumed to positively depend on two factors: the availability of an indispensable resource for both productivity and livelihood, such as water broadly conceived, and the amount of effort agents put in their productive (income-generating) actions, which itself (positively) affected by resource abundance. That is, both the resource and the appropriation effort enter in the agents' (twice-continuously differentiable) production function $f(E, R)$, where E represents the community effort (e.g. the aggregate water usage) resulting from the actions of the n agents comprising it, and R is the resource available to the community (which may either be entirely consumed in a given time period, or saved in part for future consumption). Formally, letting e_i be the individual effort (i.e. his/her resource uptake), which can either take value e_c for a cooperator or e_d for a defector, with $e_c < e_d$ due to the more sustainable practices of the former, the following inequalities are therefore assumed to hold: $\frac{\partial e_i}{\partial R} > 0$, $\frac{\partial f(E,R)}{\partial E} > 0$, $\frac{\partial f(E,R)}{\partial R} > 0$, $\frac{\partial^2 f(E,R)}{\partial E^2} \leq 0$
 $\frac{\partial^2 f(E,R)}{\partial E \partial R} \geq 0$, $\frac{\partial(f(E,R)/E)}{\partial R} \geq 0$.⁴

Let's go in further detail about the model. It is useful to consider again the joint level of effort E resulting from the actions of the n agents choosing their level of effort e_i ; letting $f_c \in [0,1]$ be the proportion of cooperators, we have $E(f_c, R) = f_c * n * e_c(R) + (1 - f_c) * n * e_d(R)$. We assume that n is fixed, so that entry is ruled out, while f_c is continuous and non-negative, and see that for positive levels of e_c and e_d , the total level of effort is a decreasing function of the frequency of cooperators.

The two effort levels, that are here assumed to sum up the behavioral inclinations of all agents in the community, are bounded below by the collectively efficient resource use level and above by the static Nash equilibrium level. This amounts to require that both agent types follow practices that are above those that would maximize collective utility, but to a different extent: the defectors ignore the emergent social norm prescribing the socially agreed-upon acceptable individual effort e_c by choosing a greater level e_d (up to the individually rational but inefficient Nash equilibrium level resulting in excessive resource use), while cooperators stick to e_c , which, as a special case, may coincide with the level that efficiently trades off the individual incentive towards high or uncoordinated resource utilization with the social need to impose constraints to guarantee a sustainable use (which ultimately benefits the individuals). Letting E_{eff} be the community efficient level, $e_{eff} = \frac{E_{eff}}{n}$ the corresponding individual efficient level under symmetry, and e_{nash} be the Nash equilibrium level of effort, we formalize what

⁴ These assumptions are generally employed in the literature concerning resource exploitation in a common pool resource, such as, for example, a fishery where a community of fishermen have access to it and each can decide on the individual level of exploitation (jointly affecting the sustainability of the resource utilization). Cf. Sethi, Somanathan (1996), Xepapadeas (2005) and Osés-Eraso, Viladrich-Grau (2007).

stated above as: $e_{eff} \leq e_c < e_d \leq e_{nash}$.⁵ These conditions guarantee that, at the aggregate level, positive rents from productive use of the resource can be maintained. That is, the average product of labor is assured to be above the opportunity cost of labor independently from the share of defectors, providing the incentive for agents to increase their resource use (as they can earn positive profits for positive levels of effort). It is further assumed that $f(0, R) = 0 = f(E, 0)$ for the obvious reason that strictly positive levels of effort and resource are required to generate income via the function $f(E, R)$.

The individual payoff given resource R and the behaviour of all community members is:

$$\pi_i(e_1, e_2, \dots, e_n, R) = p \frac{e_i}{E} f(E, R) - w e_i \quad (1)$$

Letting R^* the equilibrium resource level (to be defined more precisely below) and $\Pi(e_1, e_2, \dots, e_n, R) = \sum_{i=1}^n \pi_i = f(E, R) - wE$, the optimal solution to the aggregate payoff maximization problem is given by $E_{eff} = \arg \max_E (\Pi)$, and satisfies $f'(E_{eff}, R^*) = \frac{w}{p}$, where $\frac{w}{p}$ is the ratio between the opportunity cost of labor w and the world price p of the resource-absorbing good produced.⁶ For the sake of compactness and to stress that agent i 's payoff is only indirectly affected by the others' choice of effort (through $f(E, R)$), we will use the notation $\pi_i(e_i, R)$ below.

We know that, due to the negative appropriation externality arising from the disconnect between individual extraction and knowledge of its effect on the resource stock, the aggregate level of effort in equilibrium if all agents play according to the Nash equilibrium will be above E_{eff} as each individual will treat the resource stock as fixed and therefore extract more resource than is efficient.

Resource dynamics

Let's turn to the resource dynamics and its interaction with the social dynamics occurring as a result of human action. Focusing on the ecological features governing the resource first, in the absence of human appropriation we are left with the constant resource inflow (c) and a term dependent on the resource level (R) as well as on three parameters (d , k and R_{max} governing the discharge, curvature and maximum storage capacity) to account for a positive rate of growth up to the R_{max} (e.g. the upper limit in a groundwater aquifer's intake), which becomes negative beyond that level. We follow Ibañez (2004) for what concerns the above mentioned ecological variables, and include the aggregate resource use by the individuals (ER), which appear as the last term of the following equation:

$$\dot{R} = c - d \left(\frac{R}{R_{max}} \right)^k - ER \quad (2)$$

\dot{R} indicates the time derivative of the resource stock, i.e. its overall rate of change resulting from the interaction of replenishment, discharge and utilization. Note that the resource replenishment rate, which is captured by the first term in the right-hand side of (2), is exogenous with respect to the frequency of

⁵ Note that a direct implication of such constraints is that $E_{eff} \leq E \leq E_{nash}$. See Dasgupta and Heal (1979, 55-60) for a comprehensive treatment of exhaustible resources, and Oses-Eraso, Viladrich-Grau (2007, 398) for the description of a process leading to the prevalence of one representative strategy for each type of behavior.

⁶ An example of a function guaranteeing the existence of an optimal solution, at each point in time, to the aggregate payoff maximization problem, is the familiar Cobb-Douglas formulation with decreasing returns to scale: $f(E, R) = E_t^\alpha * R_t^\beta$, $\forall E \geq 0, R > 0$ and $\alpha + \beta < 1$.

cooperators (and consequently the resource extraction), which affects instead the second and last terms, respectively representing the limits to resource accumulation (due to stock effects) and the resource utilized by the community for productive tasks such as irrigation.

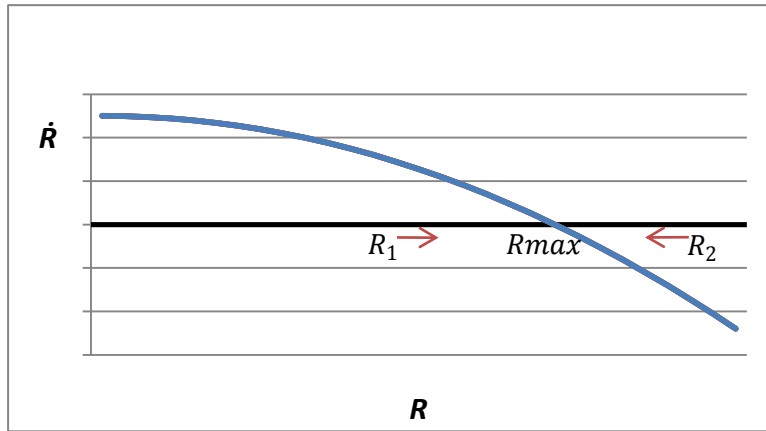


Figure 2.1: The rate of change of the resource as a function of its stock, in the absence of appropriation

For the sake of concreteness, one can think of agents extracting water for irrigation purposes from an underground reservoir subject to replenishment due to snowmelt or rain, whose ability to store water has a natural upper bound (R_{max}). Beyond it, discharge occurs at a rate which is increasing in the deviation from the maximum storage capacity. Two facts are worth noting at this point. First, in the absence of extraction, the equilibrium resource level will settle on R_{max} . This since, if the stock at one point in time is to its left (R_1 in Figure 2.1), the resource will continue to accumulate ($\dot{R} > 0$) until R_{max} is reached; if instead the stock at a given time is to its right (R_2), discharge will bring it back to R_{max} . Secondly, due to human extraction, the equilibrium resource level R^* will actually be below the maximum storage capacity.⁷ With these notions in mind, we are now ready to shift attention to the strategies and tradeoffs faced by the two types of agents.

Equity-driven ostracism

Recall that equation (1) represents the amount an individual appropriator can make based on his/her effort and the yield, abstracting from costs originating from social stigma (and the consequential ostracism imposed by the community on defectors). This amount is proportional to the aggregate payoff (itself a function of the market price of the final yield p and of the production function f), in relation to the individual's resource uptake e_i , which positively enters in the first term in the right hand side of (1) and negatively in the second term representing the work-related costs.

⁷ In fact, due to the boundary conditions on the effort levels and (1), the equilibrium resource level will satisfy the condition $0 < R_{nash} \leq R^* \leq R_{eff} < R_{max}$, where R_{nash} is the resource level corresponding to monomorphic Nash behavior (a population comprised solely of individuals maximizing profits taking R as exogenous), and R_{eff} is the socially optimal level that would obtain if all individuals jointly maximized collective welfare (effectively internalizing the appropriation externality). Therefore, depending on the population composition, and consequently on the aggregate extraction, the equilibrium level R^* will be closer to one of the above two boundaries: in a polymorphic population comprised of a majority of defectors, R^* will be close to R_{nash} , while its distance from R_{eff} will be shorter the more the number of cooperators. Note that according to the above inequality, even under full defection there is a positive resource value ($0 < R_{nash}$) guaranteeing the assumed positive rents.

To account for the costs accruing to norm violators when identified by the community as such, we incorporate them in the:

$$\mathcal{U}_i = \pi_i - \omega(f_c) * \max\left\{\frac{\pi_i - \pi_c}{\pi_d}, 0\right\} = e_i \left(p \frac{f(E,R)}{E} - w\right) - \omega(f_c) * \max\left\{\left(\frac{e_i - e_c}{e_d}\right) \left(p \frac{f(E,R)}{E} - w\right), 0\right\} \quad (3)$$

This translates to a payoff to a norm complier (cooperator) which is simply given by

$$\mathcal{U}_c = \pi_c \quad (4)$$

while a norm violator will be subject to:

$$\mathcal{U}_d = \pi_d - \omega(f_c) \left(\frac{\pi_d - \pi_c}{\pi_d}\right) \quad (5)^8$$

Recalling the tradeoffs highlighted in Figure 1, one sees from the comparison of (4) and (5) that, for what concerns the productivity, defectors have an advantage ($\pi_d > \pi_c$) as a consequence of their higher appropriation; due to stock effects, such productivity advantage positively depends on the relative abundance of cooperators. On the other hand, defectors are deprived of the communitarian social capital and experience a reduction to the income generated with resource-intensive productive activities, while cooperators can also tap in the community for help and thus enjoy the entire yield π_c . As compliance to the social norm is voluntary and observable, these benefits are denied to non-members; further, it is assumed that the community ostracism function $\omega(f_c)$ is increasing and concave in the number of participants (abiding to the social norm), due to the network good characteristics highlighted at the beginning of this section. Notice also that it multiplies the ratio between the payoff difference and the defector's payoff, to model a reaction by the cooperators which is stronger the larger the relative intensity of the defection (leading to a larger negative productivity gap of norm followers with respect to defectors).

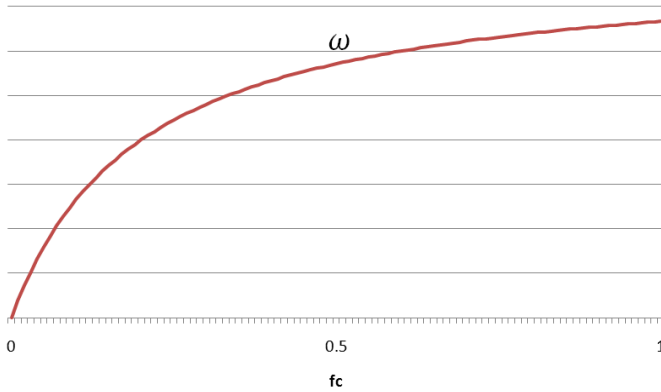


Figure 2.2: The ostracism function. A higher f_c (larger cooperator community) increases the ostracism's severity

⁸ Raakjær Nielsen and Mathiesen (2003) found, among the five more relevant factors affecting compliance in Danish fisheries: the economic gains to be obtained from noncompliance, deterrence and sanctioning costs, and the presence of “norms (behaviour of other fishers) and morals”. In (5), the gains appear in π_d , while the losses due to the sanctions and the comparison with others are captured by the product $\omega(f_c) \left(\frac{\pi_d - \pi_c}{\pi_d}\right)$. Notice that the strongest action against defectors will be taken when the number of cooperators is largest (i.e. $\pi_d - \pi_c$ and $\omega(f_c)$ are highest), while when defection is spread all over it will go almost unnoticed (i.e. $\pi_d \cong \pi_c$ and $\omega(f_c)$ is low).

Social dynamics

The analysis of the behavioral evolution of agents facing decisions on their resource practices is conducted by means of replicator dynamics. By so doing, we avoid the complete rationality requirements typical of models of optimization, while retaining (myopic and lagged) convergence towards better outcomes due to the imitation of successful behavior. Such an approach is particularly well-suited to the analysis of the evolution of norm adoptions as it allows focus on emergent phenomena without being confined, as is the case for neoclassical analysis, to equilibrium outcomes and representative agents solely described by their optimizing behavior. Here, rather than rationally best responding to the actions of others as in Nash equilibrium, individuals update their strategies when given the option, and switch to the strategy of the agent with which they are randomly matched if the utility of the latter is above the individual's⁹. It can be shown that such strategy revision takes place with a probability which is proportional to the payoff difference with respect to the average: if, for example, the average is well above the payoff of a cooperator, he or she is more likely to notice the need to switch than if the average was only slightly above the agent's payoff.

Formally, this leads to the 2-strategy replicator dynamics, which combined with (3) yields, after rearranging terms:

$$\dot{f}_c = f_c(U_c - \bar{U}) = f_c(1 - f_c)(U_c - U_d) = f_c(1 - f_c) \frac{\pi_d - \pi_c}{\pi_d} (\omega(f_c) - \pi_d) \quad (6)$$

The dotted superscript stands for time derivative: equation (6) models the evolution of cooperating types. We are interested in the nullclines satisfying $\dot{f}_c = 0$: in addition to the monomorphic outcomes characterized by one type of agent only, we look for solutions in which positive amount of both types coexist (with $f_c \neq 0$ and $f_c \neq 1$). That is,

$$(f_c^*, R^*): \theta(f_c^*, R^*) = \frac{\pi_d(e_d, R^*) - \pi_c(e_c, R^*)}{\pi_d(e_d, R^*)} (\omega(f_c^*) - \pi_d(e_d, R^*)) = 0 \quad (7)$$

The system described in (2)-(6) can be represented in the (μ, f_c) parameter space, where μ is the effort multiplier between a cooperator and a defector (e.g. $\mu = 2$ signifies that the defector's effort is twice the cooperator's effort, while μ_{nash} is consistent with $e_d = e_{nash}$).

⁹ See Taylor and Jonker (1978) or Weibull (1995) for details on replicator dynamics.

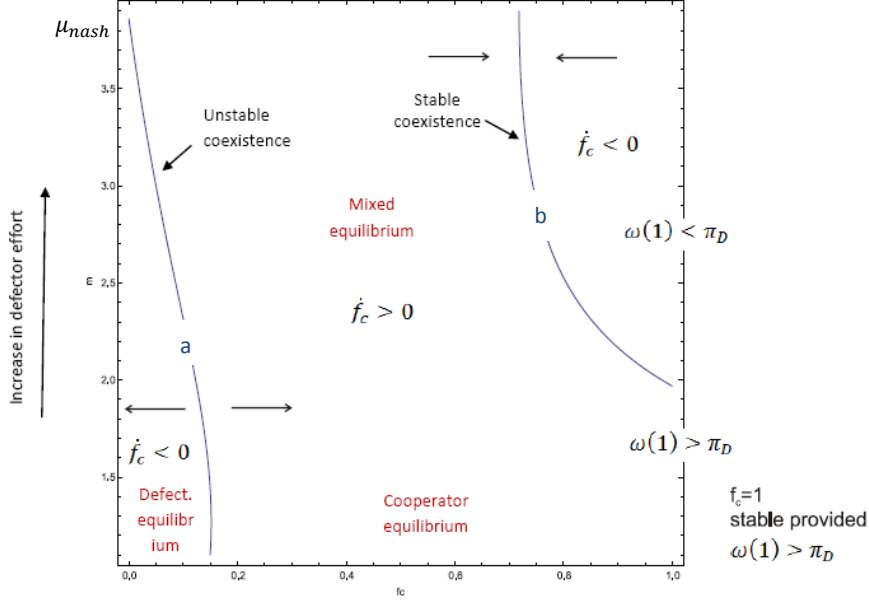


Figure 2.3: The $\omega(f_c^*) = \pi_d(e_d, R^*)$ loci guaranteeing coexistence of types

Given the positive value of the first three terms on the right hand side of (5), (with the exception of degenerate cases), it is straightforward that the proportion of cooperators will increase ($\dot{f}_c > 0$) provided that $\omega(f_c) > \pi_d(e_d, R)$. Where the system ultimately stabilizes depends on the initial conditions: however, as we can readily observe from Figure 2.3, we have both areas characterized by the presence of one type of agents only and areas of coexistence. Note that with an increase in cooperators the ostracism $\omega(f_c)$ increases, but so does the defector payoff $\pi_d(e_d, R)$ (because productivity increases with an increase in f_c); such interaction causes non-trivial dynamics that are analyzed here under the evolutionary lens provided by the replicator equation. Below we summarize the results of the stability analysis based on the linearization matrix J of the system (2)-(6):

$$J = \begin{bmatrix} \partial \frac{\dot{f}_c}{f_c} & \partial \frac{\dot{f}_c}{R} \\ \partial \frac{\dot{R}}{f_c} & \partial \frac{\dot{R}}{R} \end{bmatrix} = \begin{bmatrix} \varphi[\omega'(f_c)(f_c - f_c^2) + \omega(f_c)(1 - 2f_c) + (2f_c - 1)\pi_d] & 0 \\ nR(e_d - e_c) & -\frac{dkR^{k-1}}{R_{max}^k} - E \end{bmatrix}$$

Where $\varphi = \frac{\pi_d(e_d, R) - \pi_c(e_c, R)}{\pi_d(e_d, R)}$.

The bottom two entries of J are unambiguous in sign, which is respectively positive and negative. Therefore, for stability considerations, whether an equilibrium is stable depends on $\partial \frac{\dot{f}_c}{f_c}$. When $f_c = 0$, it reduces to $-\varphi\pi_d < 0$, so $tr(J) < 0$ and $Det(J) > 0$, which means both eigenvalues of J are negative real numbers and the all-defectors equilibrium is a stable attracting fixed point. It can be shown that, when $f_c = 1$, $\partial \frac{\dot{f}_c}{f_c} = \varphi(\pi_d - \omega(1))$ which is negative provided that $\omega(1) > \pi_d$. Thus, the all-cooperators equilibrium is stable so long as the defector's payoff is bounded above by the full compliance ostracism costs.

What happens for intermediate frequencies of cooperators? From (7) we know that, if it exists, the coexistence equilibrium must satisfy $\omega(f_c^*) = \pi_d(e_d, R^*)$. Inspection of the blue curves in Figure 2.3 allows one to assess the qualitative features of the system resulting from the above condition: to the left of locus a , i.e. for low initial f_c , $\omega(f_c) < \pi_d(e_d, R)$, so the system will evolve towards the stable defector equilibrium independently of μ . If, for instance, we consider defectors who extract resource according to the Nash rule ($\mu_{nash}: e_d = e_{nash}$), the equilibrium will be characterized by $\omega(0) = 0 < \pi_d(e_d, R_{nash})$ (see footnote 7). To the right of locus a , $\omega(f_c) > \pi_d(e_d, R)$, so the community of appropriators following the restrictive norm will grow larger. The system will transition towards the cooperator equilibrium when the effort difference between cooperators and defectors is not too large (low μ), as the above inequality will continue to hold until stable monomorphic cooperation obtains, with $\omega(1) > \pi_d(e_d, R_{eff})$ (see Figure 2.3 and footnote 7). When instead effort differences are large, the proportion of cooperators will keep increasing up to a point where $\omega(f_c^*) = \pi_d(e_d, R^*)$: At this point population composition does not change any longer and the mixed equilibrium persists. The same obtains when starting to the right of locus b ; in other words, b is a stable locus of mixed equilibria. Note that this is not true for locus a , which is unstable.

3. Model extensions with variable resource replenishment rate and strategy competition

To extend the above analysis to conditions of variable resource inflow or multiple extraction strategies we developed agent-based simulations (ABM). The basic setup of the ABM closely follows the evolutionary game theoretic model in discrete time, with individual agents being randomly matched for payoff comparison and strategy updating as in the replicator dynamics above. A time step equals the length of one replenishment cycle of the resource.

Agents extract resources according to their effort strategy and produce a good that provides them with the respective payoff. At each time step two agents are matched randomly to update their utilities and strategies. When a positive payoff difference signals defection the defector will be ostracized with a magnitude proportional to the share of cooperators in the population and the relative payoff difference between the two opponents (see equation 3). Subsequently, when his utility is below that of the opponent, an agent i updates his strategy by imitating the strategy of the opponent j with a probability equal to the utility difference (cf. Morgan 2003).

$$\Delta_i = u_i - u_j \quad \text{if } \Delta_i < 0 \rightarrow e_i = e_j \quad \text{with probability } \frac{u_i - u_j}{|u_i| + |u_j|} \quad (8)$$

We simulate the evolution of the population from an initial population composition and a specified effort difference between cooperators and defectors. Figure 3.1 shows the result of multiple simulation runs with different initial conditions and effort differences. Note that the effort difference represented on the y-axis is fixed within each simulation, while the initial proportion of cooperators represented on the x-axis is only the initial value with the final value presented in the pixels in the graph. Each pixel in the parameter space represents the average proportion of cooperators of the last 500 time steps of each run across 30 runs. With low initial proportion of cooperators and low effort difference the system

converges to a defector equilibrium (henceforth, All-D). A cooperator equilibrium (All-C) emerges when initial proportions of cooperators are at least 15% and the effort difference not too large. Once effort differences increase above approximately 2 (i.e. defector effort is twice the cooperator effort) the system converges to a mixed equilibrium. When effort differences increase even more the mixed equilibrium covers the entire parameter range and the frequency of cooperators in the mixed equilibrium decreases.

The simulation results are consistent with the analytical model presented in Figure 2.3 except for the region in the upper left corner, where in the simulations the stable defector boundary equilibrium disappears. Here, the transient dynamics characterized by stochastic updating and the discreteness of the individual agents create a situation where the resource is depleted faster than the single cooperator has a chance to update his strategy to defection. In the game theoretic model with high effort differences the unstable boundary to the mixed equilibrium approaches the stable all defector equilibrium and thus small perturbations e.g. by the stochastic updating can shift the system to the mixed equilibrium.

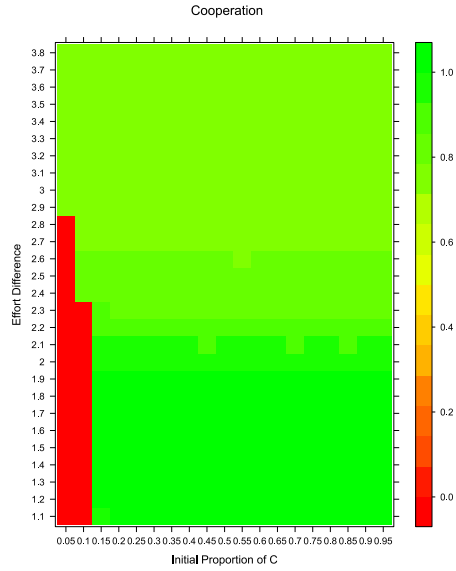


Figure 3.1: Frequency of cooperators for different initial f_c and different μ . Red indicates All-D, green All-C. Each square represents the average of 10 different runs.

Variable Inflow

In reality resource flows such as water flows in a landscape are rarely constant. Particularly in semi-arid regions water availability can vary drastically within and between years. Climate change is likely to increase this variability and lead to more frequent extreme events. This puts additional pressure on water users that have to cope and adapt to changing resource conditions. To assess the effect of inflow variability on the stability of cooperation we developed an ABM with variable inflow to the resource pool. The modified resource dynamics are given in the following relationship:

$$R_{var,t+1} = R_{var,t} + c_{var,t+1} - d_{var,t+1} * \left(\frac{R_t}{R_{max}}\right)^2 - E_t R_t \quad (9)$$

where $c_{var,t}$ (for $t = 1, \dots$) are independently distributed pseudorandom Gaussian values standardized to have mean c and standard deviation σ . The baseline discharge rate $d_{var,t}$ is set equal to the inflow to maintain a carrying capacity equal to R_{max} .

Figure 3.2 shows that with an increase in variability of inflow to the common resource pool, the percentage of cooperators in the mixed equilibrium increases, thus indicating an advantage for cooperators. We explain the disadvantage for the defectors with the concavity of the resource function at $\tilde{R} = 0$, which leads to a decrease in the average resource volume with inflow variability. Recalling the analysis of Section 2, letting $\tilde{R} < R$ be the perturbed resource volume at time t , we note that the corresponding payoff $\tilde{\pi}_i(e_i, \tilde{R}) < \pi_i(e_i, R)$, which implies that $\tilde{\pi}_d(e_d, \tilde{R}) < \pi_d(e_d, R)$. However, the ostracism function is independent of R , so the loci satisfying $\omega(f_c^*) = \pi_d(e_d, \tilde{R}^*)$ will shift rightward (a and b in Figure 2.3). Put differently, a lower average resource volume leads to reduced payoffs for both defectors and cooperators. Defectors, however, are also subject to ostracism which is only a function of the frequency of cooperators and thus is not affected by inflow variability. The decrease in defector payoff with constant ostracism costs decreases the frequency of defectors in the mixed equilibrium.

With respect to the phase plot of the game theoretic model the simulation results indicate that the loci of stable coexistence ($\omega(f_c) = \pi_d$) shift further right and closer to the unstable all cooperators boundary equilibrium.

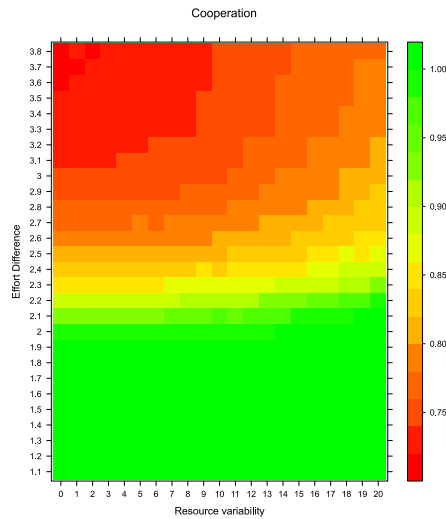


Figure 3.2. Frequency of cooperators with increasing variability of resource inflow and μ . The initial proportion of cooperators is set to 0.5. Here red indicates $f_c = 70\%$, while green indicates All-C.

Competition of Multiple Strategies

It is likely that resource users chose extraction strategies that are not confined to a set of a single cooperative and defector strategy but vary more widely. To assess how a fairness norm affects the evolution of individual extraction strategies and thus total extraction behavior of the community we developed an ABM that allows for agents with multiple strategies. Agents are initialized with a random strategy drawn from a uniform distribution on the interval $[0, \psi * E_{openaccess}/N]$, $\psi \geq 1$. Interactions between agents are modeled as before through random pairs that one at a time step update their utilities

and strategies. An agent is identified as a defector when his payoff is larger than the payoff of the opponent. Thus in this case the definition of a defector becomes dynamic and relative to the actual opponent. The magnitude of the ostracism is a function of the proportion of cooperators which is determined in the following way: cooperators are all agents who follow the norm, i.e. have an extraction strategy that is equal or lower than the socially acceptable extraction level:

$$f_c = (\sum_{j=0, j \neq i}^N n_j) / N \text{ for which } e_j \leq e_{efficient} \quad (10)$$

The ostracism is normalized using the highest payoff available in the population at the current time step, instead of the standardized defector payoff as in the game theoretic version.

$$u_{i,defector} = e_i * \left(\frac{f(E,R)}{E} - w \right) - \omega(f_c) * \frac{\max\{\pi_i - \pi_j, 0\}}{\pi_{max,t}} \quad (11)$$

We allow for random mutations where one agent chooses a new random extraction strategy at a specified mutation rate. Figures 3.3 and 3.4 show the results of the competition of multiple strategies with different strengths of ostracism expressed through the δ parameter of the ostracism function. Each point is the mean or standard deviation of 100 runs for each δ value. With increasing strength of the social norm there is a clear transition from a defector state where total effort is at open access to an all cooperator situation with effort levels at or below the socially optimal level. This transition is signaled by an increase in the standard deviation of total effort which peaks around delta values of 0.5. This is a region of bistability where some simulations result in an open access total effort, while others result in lower total extraction levels (Figure 3.3). The average number of strategies present in the population also increases during the transition and decreases significantly with an increase in delta. Thus, with increasing strength of ostracism conformity is increased.

The peak of total payoffs occurs around delta values of 1 where total effort is at the socially optimal level (0.48). With an increase in the strength of the ostracism the population converges to extraction levels that are below socially optimal indicating that the fairness norm is effective in restraining individual resource extraction but can lead to suboptimally low resource exploitation.

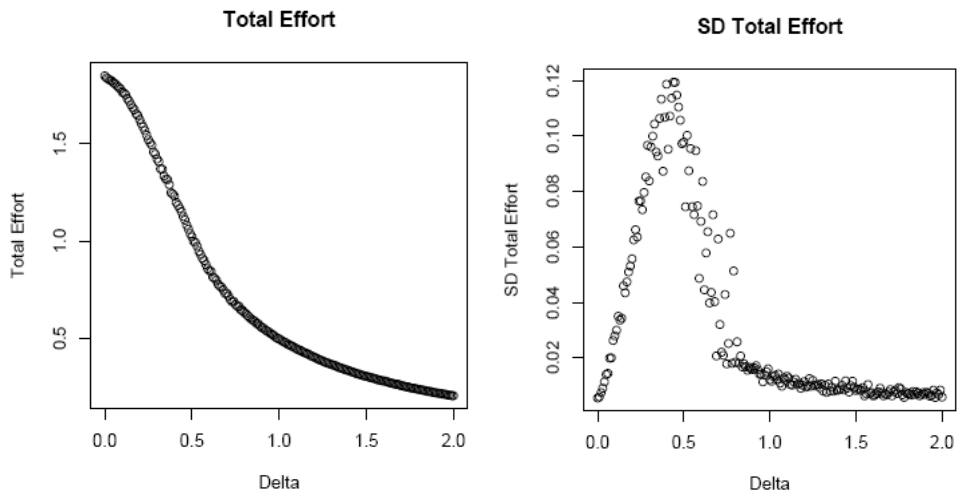


Figure 3.3: Average total effort (left) and standard deviation (right) of 50 runs for each delta value. Simulation initialized with 50 random uniformly distributed strategies. Mutation rate is 1 mutation every 10 time steps.

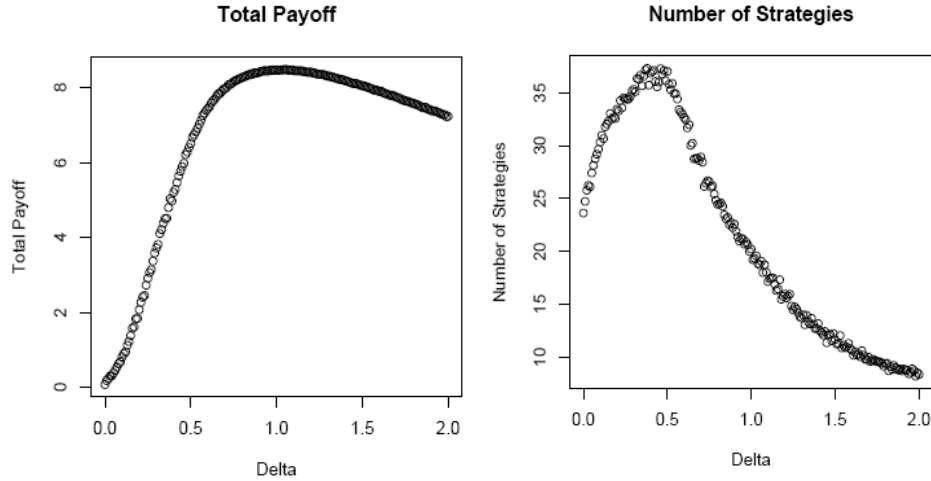


Figure 3.4: Average total effort (left) and standard deviation (right) of 50 runs for each delta value. Simulation initialized with 50 random uniformly distributed strategies. Mutation rate is 1 mutation every 10 time steps.

4. Conclusions

In this paper we have developed a model of community-based appropriation of a common pool resource in the presence of a norm allowing discrimination of behavior. Individuals departing from what the community considers as acceptable behavior (in terms of non-excessive resource extraction), are therefore subject to what we call equity-driven ostracism: a denial of support by the cooperating community which has tangible consequences on the wealth of the norm violators. Such retaliation, which may take the form of spiteful actions (e.g. denied machinery lending and crop destruction) or social reprobation (e.g. negative gossiping and refusal to share information), depends on two factors. On the one hand, the relative strength of the community of norm followers, since a larger community is assumed to be more effective at ostracizing defectors. Secondly, the intensity of the response by the community is assumed to be higher the larger the entity of the defection, which is revealed by the yield of the production as it depends on the amount of resource extraction. Unlike Osés-Eraso, Viladrich-Grau (2007), we don't have to make as strong an assumption as in their footnote 9, namely of perfect observability of effort. Our more nuanced version requires rather perfect observability of the outcome of the production. To our knowledge, this is the first effort aimed at modeling the coupled socio-economic and ecological dynamics of a common pool resource, such as water, that provides benefits indirectly by being utilized as an input of production rather than for its intrinsic value (as is the case for fish in fisheries).

To recapitulate, the model presented here provides insights into the evolution of cooperation among agents deciding their patterns of extraction of a shared resource in the presence of a social norm discerning between acceptable from excessive behavior. Analytical derivations and evidence from numerous simulations allowing for complex interactions and resource dynamics lead to the following conclusions:

- I. All-D is achieved only for low initial frequency of cooperators; interestingly, the basin of attraction shrinks as the gap between e_d and e_c (i.e. μ) increases, due to resource stock effects. All other configurations lead either II or III, even when most agents initially do not abide to the norm.
- II. All-C arises so long as the defector's payoff is bounded above by the full compliance ostracism costs: $\omega(1) > \pi_d(e_d, R_{eff})$. This is the case when the effort differences between cooperators and defectors are not too large.
- III. Stable coexistence obtains when effort differences between cooperators and defectors are pronounced, sparing the latter from being eradicated: $\omega(f_c^*) = \pi_d(e_d, R^*)$
- IV. Under variable resource replenishment rates, cooperators thrive better, because they can still benefit from the social capital provided by other cooperators despite a reduction in average resource volumes, while the defectors experience a decrease in payoffs to $\tilde{\pi}_d(e_d, \tilde{R})$.
- V. A competition of extraction strategies will lead to open access extraction efforts when the social norm is weak; however the system will approach optimal extraction levels rapidly when the strength of the social norm increases.

We are currently working on increasing the degree of realism of the model to account for other relevant features concerning both the social and the ecological dimensions. In particular, the present work will benefit from allowing for agent interaction through social networks and the introduction of a gradient in resource access. Moreover, we plan to perform a series of lab and field experiments in Central Asia to gain insights on these complex dynamics in the presence of severe climatic conditions and resource scarcity.

5. References

Chen, X. et al.: Linking social norms to efficient conservation investment in payments for ecosystem services. *Proceedings of the National Academy of Sciences*, 106(28) (2009)

Dasupta, P., Heal, G.: *Economic theory and exhaustible resources*. Cambridge University Press, Cambridge (1979)

Ibáñez, J., Martínez, S., Martínez, J.: Competitive and optimal control strategies for groundwater pumping by agricultural production units. *Water Resources Research* 40 (2004)

Levin, S.: Learning to live in a global commons: socioeconomic challenges for a sustainable environment. *Ecological Research* 21(3) (2006)

Lindbeck, A.: Incentives and Social Norms in Household Behavior. *The American Economic Review* 87(2) (1997)

Maier-Rigaud, F.P., Martinsson, P., Staffiero, G., Ostracism and the Provision of a Public Good Experimental Evidence. *Journal of Economic Behavior and Organization* (2008)

Morgan, J., Steiglitz, K.: Pairwise competition and the replicator equation. *Bulletin of Mathematical Biology* 65(6) (2003)

Noailly, J., Withagen, C., van den Bergh, J.: Spatial Evolution of Social Norms in a Common-Pool Resource Game. *Environmental & Resource Economics* 36(1) (2007)

Nuria Osès-Eraso, N., Montserrat, V.-G.: On the sustainability of common property resources. *Journal of Environmental Economics and Management* 53 (2007)

Ostrom, E.: *Governing the Commons: The Evolution of Institutions for Collective Action* Ostrom, Elinor, Cambridge University Press (1990)

Ostrom, E.: Challenges and growth: the development of the interdisciplinary field of institutional analysis. *Journal of Institutional Economics* 3(03) (2007)

Sethi, R., Somanathan, E.: The Evolution of Social Norms in Common Property Resource Use. *The American Economic Review* 86(4) (1996)

Raakjær Nielsen, J., Mathiesen, C.: Important factors influencing rule compliance in fisheries lessons from Denmark. *Marine Policy* 27(5) (2003)

Taylor, P., Jonker, L.: Evolutionarily Stable Strategies and Game Dynamics. *Mathematical Biosciences* 40(2) (1978)

Xepapadeas, A.: Regulation and Evolution of Compliance in Common Pool Resources. *The Scandinavian Journal of Economics* 107(3) (2005)