Guards vs Vigilantes: The Effect of Rule Enforcement Strategies on Sustainable Use Norms in Common Property Regimes

Trevor Breen[i] and Alessandro Tavoni[ii]

Abstract

This paper uses replicator dynamics to compare the steady states arising from two types of common property regimes - one in which over-exploiters are punished by the resource users themselves, and another where enforcement is handled by guards who collect a tax from the users. The use of guards requires a less restrictive set of parametric conditions in order to maintain an equilibrium with no over-exploiters. However, it can also stabilize an outcome in which all users over-exploit and are punished, but not enough to induce more cooperation (less resource extraction). These results can be used in guiding and structuring the formation of new common property regimes.

JEL Codes: *C7, Q2, Q5*

Keywords: *Common Property, Evolutionary Game Theory, Institutions, Punishment*

[i]University of Alberta, Canada T6G 2R3 and Grantham Research Institute on Climate Change and the Environment, London School of Economics, U.K., WC2A2AE. tbreen@ualberta.ca

[ii]Corresponding Author, Department of Economics, University of Bologna, Italy, 40126 and Grantham Research Institute on Climate Change and the Environment, London School of Economics, U.K., WC2A2AE. alessandro.tavoni2@unibo.it

1 Introduction

Neoclassical economic models of resource use tend to make grim predictions regarding the viability of common property regimes (CPRs), in which the rights to access a resource are shared by a defined group of users. In such an arrangement, the Nash equilibrium level of effort will be strictly greater than the socially optimum for any population greater than 1 (Hamburger 1973). If punishment for over-extraction could be effectively enforced in a CPR, and the cost of being punished was sufficiently high, then in principle the socially optimal level of extraction could be induced. However, monitoring and enforcing rules around resource use is typically costly, so there is an incentive for individuals to free-ride on punishment efforts undertaken by others. Consequently, rationally self-interested agents would not be able to self-enforce extraction limits in CPRs (Henrich and Boyd 2001, Fehr and Gächter 2002).

Stock dependent growth effects tend to create even stronger incentives towards over-extraction - since an individual in a CPR with n users appropriates only a $1/n$ share of the stock, s/he only internalizes a $1/n$ share of the reduction in stock-growth that results from over-extraction (Dasgupta and Heal 1979, Perman 2003). This would lead us to expect unsustainable extraction within CPRs tending towards stock collapse. These effects are attenuated when the population of users is small and fixed, but as Hardin (1968) notes, population growth causes CPRs to approach open access conditions even when the resource is controlled by a single group of users.

While there are examples of failed CPRs, there have been many examples of successful CPRs with active mechanisms to enforce harvest restrictions (see Baland and Platteau 1996 and Ostrom 1990 for a review). These examples defy the reasoning presented above. One potential explanation for this discrepancy is the strong information assumptions in the models used by Perman (2003) and Dasgupta and Heal (1979) - they assume perfect information about the payoff function, universal rationality, and common knowledge about that rationality.[3] Uncertainty about the payoff functions or about the reasoning of other players can create the incentive to use an 'imitate the successful' heuristic rather than Nash equilibrium reasoning (Camerer 1997, Gigerenzer and Selten 2002).

Sethi and Somanathan (1996) present a model which captures these insights particularly well. The authors suggest that individuals may be partly motivated to cooperate by social norms, and use evolutionary dynamics to explain how such norms could emerge. Their key result is that if the cost of being punished is sufficiently high relative to the benefits of over-extraction, and the number of users engaging in over-extraction is low relative to the number of people punishing cheaters, then the system will evolve towards a cooperative outcome where no one over-extracts the resource.[4] The main contribution of the present paper is to assess the prospects of cooperation in a CPR when punishment of defectors is delegated to specialized guards who, unlike in Sethi and Somanathan (1996), do not rely on harvesting for income, but rely on taxes collected from resource users instead.

1.1 Methods

Suppose a group of n users faces some harvest function $H(X)$, which defines the total returns of a resource given a level of total harvest effort, and where each chooses some effort x_i so that $X = \sum_{i=0}^{n} x_i$. We will assume that the harvest function is concave and with diminishing returns to scale ($H' < 0$, $H'' > 0$), that no returns can be realized without effort ($H(0) = 0$), and that some returns can be profitably extracted given some fixed effort cost w, ($H'(0) > w$). The average return to effort is assumed to be constant for all users given the total group effort and defined as follows: $A(X) = H(X)/X$.

Following Sethi and Somanathan (1996), we assume that the community is composed of three types of players - cooperators (c), who contribute x_c, defectors (d), who contribute x_d and face some punishment δ from each enforcer, and enforcers (e), who contribute x_c and undertake costly punishment against each defector at a cost of γ each. Since the model assumes that defectors are punished by fellow resource users, we will refer to this setup as the *vigilante model*. The relative

[3] For challenges to these assumptions see Schultz et al. (2007), and Cialdini et al. (2004).

[4] Modifications of Sethi and Somanathan (1996) have rationalized empirically observed situations of stable coexistence between high- and low- extractors, resulting in partial internalization of the externality: Noailly et al. (2007) and Schlüter et al. (2016) relax the mean-field approximation by modelling interactions in space, while Tavoni et al. (2012) and Lade et al. (2013) introduce ostracism of defectors who exceed the norm about sustainable harvesting and find regions of coexistence of both types. A related setup is analysed in Andreoni and Gee (2011).

shares of each type sums to 1 and are governed by a replicator dynamic $\dot{s_j} = s_j(\pi_j - \bar{\pi})$ where $\bar{\pi}$ represents the average payoff in the group $(s_c\pi_c + s_d\pi_d + s_e\pi_e)$. The payoffs for each type are as follows:

$$\bar{x} = (x_c(1 - s_d) + x_d s_d) \tag{1}$$
$$\pi_c = x_c(A(n\bar{x}) - w) \tag{2}$$
$$\pi_d = \frac{x_d}{x_c}\pi_c - n\delta s_e \tag{3}$$
$$\pi_e = \pi_c - n\gamma s_d \tag{4}$$
$$\bar{\pi} = \pi_c(\frac{\bar{x}}{x_c}) - s_d s_e(\delta + \gamma) \tag{5}$$
$$\tag{6}$$

Enforce is weakly dominated by cooperate, so $s_e = 0$ in a classical game theoretic set up. This condition implies that defect weakly dominates cooperate and $s_d = 1$ (defect is the Nash Equilibrium solution). In fact, $s_d = 1$ is an unconditionally stable steady state of this system, representing a breakdown of cooperation where all users harvest unsustainably. However, $s_d = 0$ is conditionally stable, provided $n\delta > (x_c - x_d)(A(x_c) - w)$. This result means that if the number of users who are enforcing the rules is sufficiently high in the initial period, and the cost of punishment is sufficiently low, then the number of defectors will decay more quickly than the number of enforcers, and enough enforcers will remain to deter any future defectors.

CPRs enforced by vigilante justice similar to this setup include the famous lobster fisheries of Maine (Acheson 1975), as well as the coastal fisheries in the Bahia province of Brazil (Cordell and Mckean 1986). However, many of the CPRs documented in the empirical literature exhibit a more organized system in which guards are paid to enforce resource use restrictions using some form of taxation levied on the users. Such systems have been used to manage mountain pastures and forests in the Italian Alps (Casari 2007), as well as traditional common lands in medieval Japan (Mckean 1992), and a mountain community in Törbel, Switzerland (Netting 1976). Similar common property forests also existed in the Indian province of Andhra Pradesh and the Kumaon region (Wade 1989, Agrawal 2001). Does such a setup offer any advantages over informal enforcement? By modifying the payoff functions to more accurately represent such a system we can address this question. We will refer to this modified setup as the *guard model.* Given some tax rate, α, we can represent this arrangement with the following payoff functions. Note that the level of effort now depends on both s_c and s_d, since enforcers no longer harvest resources.

$$X^* = n(s_d x_d + s_c x_c) \tag{7}$$
$$\pi_c^* = x_c(\frac{H(X^*)}{X^*} - w)(1 - \alpha) \tag{8}$$
$$\pi_d^* = \frac{x_d}{x_c}\pi_c^* - n\delta s_e \tag{9}$$
$$\pi_e^* = \frac{\alpha}{s_e(1 - \alpha)}(s_c\pi_c^* + s_d\frac{x_d}{x_c}\pi_c^*) - n\gamma s_d \tag{10}$$
$$\bar{\pi}^* = \pi_c^*(s_c + \frac{x_d}{x_c}s_d) - ns_e s_d(\gamma + \delta) \tag{11}$$
$$\tag{12}$$

We can construct the Jacobian matrix for the system by taking all partial derivatives. The general form of this matrix is as follows (the guard model uses π^* in place of π):

$$\begin{bmatrix} \pi_c - \bar{\pi} + s_c(\frac{\delta(\pi_c - \bar{\pi})}{\delta s_c}) & s_c(\frac{\delta(\pi_c - \bar{\pi})}{\delta s_d}) \\ s_d(\frac{\delta(\pi_d - \bar{\pi})}{\delta s_c}) & \pi_d - \bar{\pi} + s_d(\frac{\delta(\pi_d - \bar{\pi})}{\delta s_d}) \end{bmatrix} \tag{13}$$

A potential steady state equilibrium will arise whenever $s_j = 0$ or $\pi_j - \bar{\pi} = 0$ for some strategy j. That equilibrium will be stable if the Jacobian has a positive determinant and a negative trace. There are seven potential steady-state conditions in either model - three steady states where only one type remains, three where one type disappears and the other two earn equal payoffs, and one where all three types earn equal payoffs.

2 Results

Table 1 compares the equilibria of the guard model with the established vigilante model.

Table 1: Steady State Conditions

Potential Equilibria	Vigilante Stability Conditions	Guard Stability Conditions
Cooperators Only	Unstable	Unstable
Defectors Only	Stable	Stable
Enforcers Only	$n\delta > (x_d - x_c)(A(X) - w)$	Unstable
No Cooperators*	Unstable	$\frac{\delta}{\gamma} > \frac{(x_d-x_c)(1-\alpha)}{\alpha x_d}$
No Defectors	$S_c < 1 - \frac{(x_d-x_c)(A(X)-w)}{n\delta}$	$S_c < 1 - \frac{(x_d-x_c)(1-\alpha)(A(X)-w)}{n\delta}$
No Enforcers**	Stable	Stable
All Payoffs Equal***	Unstable	Unstable
* *Conditionally stable only in guard model*		
** *Coincides with Defectors Only equilibria in both models*		
*** *Coincides with No Defectors in vigilante model but not guard model*		

Result 1. *A defector only outcome is stable in both scenarios.*

As expected, both models admit for the possibility of an unconditionally stable outcome of all defectors. This outcome arises when there is a breakdown in cooperation - the number of enforcers and cooperators dwindle to zero.

Result 2. *A no-defector outcome exists in both scenarios, but the guard model requires weaker parametric conditions for stability than the vigilante model.*

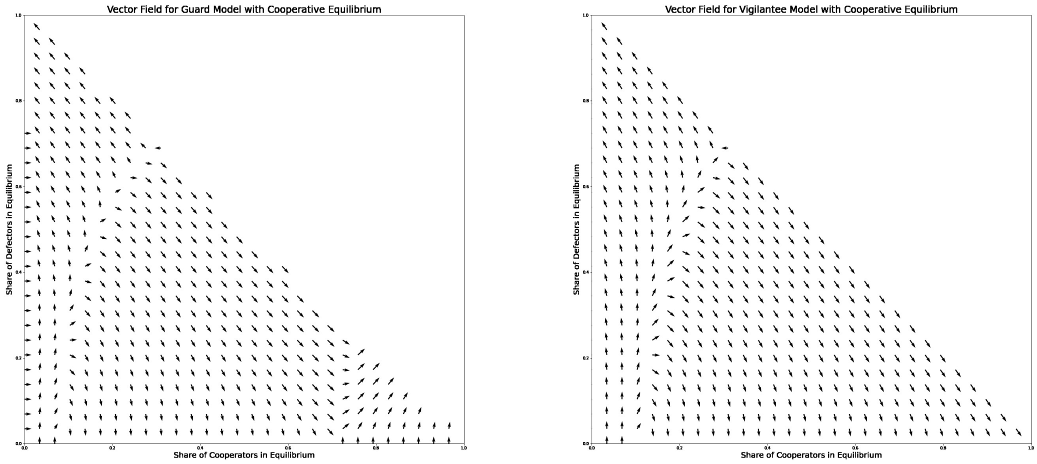
In both cases, stability depends on the ratio between the net harvest premium earned by a defector and the total punishment they receive. Specifically, a guard system will have a stable cooperative outcome if, in equilibrium, $S_c < 1 - \frac{(x_d-x_c)(1-\alpha)(A(X)-w)}{\delta}$, while under a vigilante scenario $S_c < 1 - \frac{(x_d-x_c)(A(X)-w)}{\delta}$. This result arises because the net harvest premium is attenuated by the tax charged under a guard system. Another consequence is that the stability of the cooperative equilibrium can be maintained with fewer enforcers.

This is an important result. Ciriacy-Wantrup and Bishop (1975) and Lawry (1990) identify that interactions with external economies can disrupt common property regimes. In either setup, the disruption can come as a consequence of access to new technology (increasing the average return to effort and thus the defector harvest premium), or from reductions to the damage from punishment (social sanctions such as ostracism or excommunication may be less effective if one can integrate into a sufficiently large outside group). However, this result illustrates how CPRs enforced by guards can be more resilient to such shocks.

Result 3. *An equilibrium consisting of only enforcers and defectors is possible in the guard model but not in the vigilante model.*

An equilibrium consisting of only defectors and enforcers is unstable in the vigilante model since enforcers earn a weakly lower payoff than cooperators in that system. However, it can be stable in the guard model if $\frac{\delta}{\gamma} > \frac{(x_d-x_c)}{x_d}\frac{(1-\alpha)}{\alpha}$. This is a worst-case scenario, since the resource is being used unsustainably and additional welfare is lost to enforcement that is ultimately ineffective; whether or not Cooperate will be strictly dominated depends on three ratios: the cost of being punished relative to the cost of punishing, the premium from defecting relative to the defector payoff, and the share of revenue going to harvesters relative to the share for enforcers. This new potential equalibria can be seen in the following figure, which has been parameterized to allow for the existence of a defector-enforcer equilibrium.

Basins of Attraction in Guard and Vigilantee Models



3 Discussion

3.1 Limitations

We assume that the number of enforcers could vary in response to the economic incentives at play. However, in many of the case studies cited in this paper, guards and other types of rule enforcers are hired and legitimized by a community authority that can potentially restrict the number of enforcers. Replicator dynamics are, at best, a coarse representation of the deliberative processes culminating in the choice of guards, setting of tax rates and fines.

Another stark assumption is that all individuals pay their taxes. In reality, there is a strategic incentive to avoid paying taxes (or under-report income), while free-riding on the enforcement efforts of the guards. While it is possible to imagine a simultaneous tax collection game in which the guards enforce both resource use rules and tax collection rules, such a process is outside the tractability of our simple model.

Lastly, in the guard model we assume that the cooperators and defectors 'take up the slack' whenever the share of enforcers increases, implying that at a minimum, the socially optimal effort will be contributed if $S_e < 1$. This modelling assumption implies that the cooperator-enforcer equilibrium will produce the same average payoff in both models. In reality, each person has a maximum effort level reflecting their capacity constraints. Depending on the population, the abundance of the resource stock, and the equilibrium ratio of cooperators to enforcers, the socially optimal total effort may exceed the capacity constraints of the population. At this point, the average-payoff of the vigilante model will be strictly greater than that of the guard model.

3.2 Conclusions

The results of this analysis highlight an important trade-off between formal and informal approaches to rule enforcement in CPRs - the use of formal guards can make cooperative outcomes more stable but also make it possible to realize a scenario in which the resource is used unsustainably and the community still bears the costs of rule enforcement. If a social planner found it impossible or undesirable to create formal rules guiding the use of a common property resource but was able to make certain methods of decentralized punishment possible (or salient), then these results could be used to determine which methods (and under what conditions) to make available. Given that most of the world's physical resources face some pre-existing institutions and norms, the most likely candidates for these designs would be online spaces and interactions or new off-grid communities.

4 References

References

Acheson, J. M. (1975). The lobster fiefs: economic and ecological effects of territoriality in the maine lobster industry. *Human Ecology*, *3*(3), 183–207.

Agrawal, A. (2001). Common property institutions and sustainable governance of resources. *World development*, *29*(10), 1649–1672.

Andreoni, J., & Gee, L. K. (2012). Gun for hire: Delegated enforcement and peer punishment in public goods provision. *Journal of Public Economics*, *96*(11-12), 1036–1046.

Baland, J.-M., & Platteau, J.-P. (1996). *Halting degradation of natural resources: is there a role for rural communities?* Food & Agriculture Org.

Camerer, C. F. (1997). Progress in behavioral game theory. *Journal of economic perspectives*, *11*(4), 167–188.

Casari, M. (2007). Emergence of endogenous legal institutions: Property rights and community governance in the italian alps. *Journal of Economic History*, 191–226.

Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annu. Rev. Psychol.*, *55*, 591–621.

Ciriacy-Wantrup, S. V., & Bishop, R. C. (1975). " common property" as a concept in natural resources policy. *Natural resources journal*, *15*(4), 713–727.

Cordell, J. C., & McKean, M. A. (1986). Sea tenure in bahia, brazil. In *Proceedings of the conference on common property resource management.*

Dasgupta, P. S., & Heal, G. M. (1979). *Economic theory and exhaustible resources.* Cambridge University Press.

Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, *415*(6868), 137.

Gigerenzer, G., & Selten, R. (2002). *Bounded rationality: The adaptive toolbox.* MIT press.

Hardin, G. (n.d.). The tragedy of the commons. *Science*, *162*.

Henrich, J., & Boyd, R. (2001). Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. *Journal of theoretical biology*, *208*(1), 79–89.

Lade, S. J., Tavoni, A., Levin, S. A., & Schlüter, M. (2013). Regime shifts in a social-ecological system. *Theoretical Ecology*, *6*(3), 359–372.

Lawry, S. W. (1990). Tenure policy toward common property natural resources in sub-saharan africa. *Nat. Resources J.*, *30*, 403.

Maass, A., Anderson, R. L., et al. (1978). *... and the desert shall rejoice: conflict, growth, and justice in arid environments.* MIT press.

McKean, M. A. (1992). Management of traditional common lands (iriaichi) in japan. *Making the commons work: Theory, practice, and policy.*, *339*.

Netting, R. M., & Netting, R. M. (1981). *Balancing on an alp: ecological change and continuity in a swiss mountain community.* CUP Archive.

Ostrom, E. (1990). *Governing the commons.* Cambridge university press.

Perman, R., Ma, Y., McGilvray, J., & Common, M. (2003). *Natural resource and environmental economics.* Pearson Education.

Schlüter, M., Tavoni, A., & Levin, S. (2016). Robustness of norm-driven cooperation in the commons. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20152431.

Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological science*, *18*(5), 429–434.

Sethi, R., & Somanathan, E. (1996). The evolution of social norms in common property resource use. *The American Economic Review*, 766–788.

Tavoni, A., Schlüter, M., & Levin, S. (2012). The survival of the conformist: Social pressure and renewable resource management. *Journal of Theoretical Biology*, *299*, 152–161.

Wade, R. (1989). *Village republics* (No. 40). Cambridge University Press.

5 Appendix I - Steady State Calculations

5.1 Steady State Calculations for The Guard Model

Definitions:

$$\pi_c = x_c(1-\alpha)(A(X)-w) \quad (14)$$
$$\pi_d = x_d(1-\alpha)(A(X)-w) - n\delta S_e \quad (15)$$
$$\pi_e = \frac{(S_c x_c + S_d x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma S_d \quad (16)$$
$$1 = S_c + S_d + S_e \quad (17)$$

5.1.1 Cooperators only - $S_d = 0$, $S_e = 0$, $\bar{\pi} = \pi_c \implies S_c = 1$

5.1.2 Defectors only - $S_c = 0$, $S_e = 0$, $\bar{\pi} = \pi_d \implies S_d = 1$

5.1.3 Enforcers only - $S_c = 0$, $S_d = 0$, $\bar{\pi} = \pi_c \implies S_e = 1$

5.1.4 No cooperators - $S_c = 0$, $\pi_d = \pi_e = \bar{\pi}$

From the definitions of the profit functions and the condition that $S_c = 0$ (18)

$$x_d(1-\alpha)(A(X)-w) - n\delta S_e = \frac{(0x_c + S_d x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma S_d \quad (19)$$

Factoring out S_d (20)

$$\pi_d = S_d\left(\frac{(x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma\right) \quad (21)$$

Isolating S_d (22)

$$S_d = \frac{\pi_d}{\frac{(x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma} \quad (23)$$

Applying the fact that $S_d + S_e = 1 | S_c = 0$ (24)

$$S_d = \frac{\pi_d}{(S_d + S_e)(\frac{(x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma)} \quad (25)$$

Applying the definition of enforcer profit (26)

$$S_d = \frac{\pi_d}{\pi_e + S_e(\frac{(x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma)} \quad (27)$$

Applying the condition that $\pi_d = \pi_e$ (28)

$$S_d = \frac{\pi_d}{\pi_d + x_d\alpha(A(X)-w) - S_e n\gamma} \quad (29)$$

Since S_d cannot exceed 1 (30)

$$x_d\alpha(A(X)-w) > S_e n\gamma \quad (31)$$

$$S_e < \frac{x_d\alpha(A(X)-w)}{n\gamma} \quad (32)$$

Applying the definition of π_d (33)

$$S_d = \frac{x_d(1-\alpha)(A(X)-w) - n\delta S_e}{x_d(1-\alpha)(A(X)-w) - n\delta S_e + x_d\alpha(A(X)-w) - S_e n\gamma} \quad (34)$$

$$S_d = \frac{x_d(1-\alpha)(A(X)-w) - n\delta S_e}{x_d(A(X)-w) - nS_e(\delta+\gamma)} \quad (35)$$

(36)

5.1.5 No defectors - $S_d = 0$, $\pi_c = \pi_e = \bar{\pi}$

$$\text{From the definitions of the profit functions and the condition that } S_d = 0 \tag{37}$$
$$x_c(1-\alpha)(A(X)-w) = \frac{(S_c x_c + 0(x_d))(\alpha)(A(X)-w)}{S_e} - 0(n\gamma) \tag{38}$$
$$\text{Cancelling common factors} \tag{39}$$
$$(1-\alpha) = \frac{S_c \alpha}{S_e} \tag{40}$$
$$\text{Simplifying} \tag{41}$$
$$(1-S_c)(1-\alpha) = S_c \alpha \tag{42}$$
$$1 - \alpha - S_c = 0 \tag{43}$$
$$S_c = 1 - \alpha \tag{44}$$
$$S_e = \alpha \tag{45}$$
$$\tag{46}$$

5.1.6 No enforcers - $S_e = 0$, $\pi_c = \pi_d = \bar{\pi}$

$$\text{From the definitions of the profit functions and the condition that } S_e = 0 \tag{47}$$
$$x_c(1-\alpha)(A(X)-w) = x_d(1-\alpha)(A(X)-w) - 0 \tag{48}$$
$$\text{Cancelling common factors} \tag{49}$$
$$x_c = x_d \tag{50}$$
$$\tag{51}$$

But, $x_c < x_d$ by assumption, so no such equilibrium exists.

5.1.7 All payoffs equal - $\pi_c = \pi_d = \pi_e = \bar{\pi}$

$$\text{From the definitions of the profit functions and the condition that } \pi_c = \pi_d \quad (52)$$

$$x_c(1-\alpha)(A(X)-w) = x_d(1-\alpha)(A(X)-w) - n\delta S_e \quad (53)$$

$$\text{Simplifying} \quad (54)$$

$$S_e = \frac{(x_d - x_c)(1-\alpha)(A(X)-w)}{n\delta} \quad (55)$$

$$\text{From the definitions of the profit functions and the condition that } \pi_e = \pi_c \quad (56)$$

$$\frac{(S_c x_c + S_d x_d)(\alpha)(A(X)-w)}{S_e} - n\gamma S_d = \pi_c \quad (57)$$

$$\text{Apply the definition of } S_e \quad (58)$$

$$\frac{n\delta(S_c x_c + S_d x_d)(\alpha)(A(X)-w)}{(x_d - x_c)(1-\alpha)(A(X)-w)} - n\gamma S_d = \pi_c \quad (59)$$

$$\text{Cancel Common Factors} \quad (60)$$

$$\frac{n\delta(S_c x_c + S_d x_d)(\alpha)}{(x_d - x_c)(1-\alpha)} - n\gamma S_d = \pi_c \quad (61)$$

$$\text{Isolate } S_c \text{ and } S_d \quad (62)$$

$$S_c(\frac{n\delta x_c \alpha}{(1-\alpha)(x_d - x_c)}) + S_d \frac{n\delta x_d \alpha}{(1-\alpha)(x_d - x_c)} - n\gamma S_d = \pi_c \quad (63)$$

$$\text{Express } S_c \text{ in terms of } S_d \quad (64)$$

$$S_c = \frac{(1-\alpha)(x_d - x_c)}{n\delta x_c \alpha}(\pi_c - S_d \frac{n\delta x_d \alpha}{(1-\alpha)(x_d - x_c)} + n\gamma S_d) \quad (65)$$

$$\text{Simplify} \quad (66)$$

$$S_c = \frac{\pi_c(1-\alpha)(x_d - x_c)}{n\delta x_c \alpha} + \frac{n\gamma S_d(1-\alpha)(x_d - x_c)}{n\delta x_c \alpha} - S_d \frac{(x_d)}{x_c} \quad (67)$$

$$S_c = \frac{(1-\alpha)^2(x_d - x_c)(A(X)-w)}{n\delta\alpha} + S_d(\frac{n\gamma S(1-\alpha)(x_d - x_c)}{n\delta x_c \alpha} - \frac{(x_d)}{x_c}) \quad (68)$$

$$\text{Solve for } S_d \quad (69)$$

$$S_c(\frac{n\delta x_c \alpha}{(1-\alpha)(x_d - x_c)}) + S_d(\frac{n\delta x_d \alpha}{(1-\alpha)(x_d - x_c)} - n\gamma) = \pi_c \quad (70)$$

$$S_d = (\frac{(1-\alpha)(x_d - x_c)}{n\delta x_d \alpha} - n\gamma)(x_c(A(X)-w)(1-\alpha) - S_c(\frac{n\delta x_c \alpha}{(1-\alpha)(x_d - x_c)}) \quad (71)$$

(72)

5.2 Steady State Populations for The Vigilantee Model

Definitions:

$$\pi_c = x_c(A(X) - w) \quad (73)$$

$$\pi_d = x_d(A(X) - w) - n\delta S_e \quad (74)$$

$$\pi_e = x_c(A(X) - w) - n\gamma S_d \quad (75)$$

$$1 = S_c + S_d + S_e \quad (76)$$

5.2.1 Cooperators only - $S_d = 0,\ S_e = 0,\ \bar{\pi} = \pi_c \implies S_c = 1$

5.2.2 Defectors only - $S_c = 0,\ S_e = 0,\ \bar{\pi} = \pi_d \implies S_d = 1$

5.2.3 Enforcers only - $S_c = 0,\ S_d = 0,\ \bar{\pi} = \pi_c \implies S_e = 1$

5.2.4 No cooperators - $S_c = 0,\ \pi_d = \pi_e = \bar{\pi}$

From the definitions of the profit functions and the condition that $\pi_d = \pi_e$ (77)

$$x_d(A(X) - w) - n\delta S_e = x_c(A(X) - w) - n\gamma S_d \tag{78}$$

Rearranging and simplifying (79)

$$\frac{(x_d - x_c)(A(X) - w)}{n} = \delta S_e - \gamma S_d \tag{80}$$

Apply the condition fact that $S_e = 1 - S_d | S_c = 0$ and simplify (81)

$$\frac{(x_d - x_c)(A(X) - w)}{n} = \delta(1 - S_d) - \gamma S_d \tag{82}$$

$$\frac{(x_d - x_c)(A(X) - w)}{n} - \delta = -S_d(\delta + gamma) \tag{83}$$

$$S_d = \frac{n\delta - (x_d - x_c)(A(X) - w)}{n(\delta + \gamma)} \tag{84}$$

Trivially (85)

$$S_e = 1 - S_d \tag{86}$$

$$S_e = 1 - \frac{n\delta - (x_d - x_c)(A(X) - w)}{n(\delta + \gamma)} \tag{87}$$

5.2.5 No defectors - $S_d = 0,\ \pi_c = \pi_e = \bar{\pi}$

From the definitions of the profit functions and the condition that $\pi_d = \pi_e$ (88)

$$x_c(A(X) - w) = x_c(A(X) - w) - S_d(n\gamma) \tag{89}$$

From the condition that $S_d = 0$, a tautology follows (90)

$$x_c(A(X) - w) = x_c(A(X) - w) \tag{91}$$

thus, $S_d = 0$ is sufficient for a steady state, regardless of S_c and S_e (92)

5.2.6 No enforcers - $S_e = 0,\ \pi_c = \pi_d = \bar{\pi}$

From the definitions of the profit functions and the condition that $\pi_d = \pi_e$ (93)

$$x_c(A(X) - w) = x_d(A(X) - w) - 0(n\gamma) \tag{94}$$

Cancelling out common factors (95)

$$x_d(A(X) - w) = x_c \tag{96}$$

Since cooperator and defector payoffs cannot be equal without enforcers, no such equilibrium exists.

5.2.7 All payoffs equal - $\pi_c = \pi_d = \pi_e = \bar{\pi}$

From the definitions of the profit functions and the condition that $\pi_c = \pi_e$ (97)

$$x_c(A(X) - w) = x_c(A(X) - w) - S_d(n\gamma) \tag{98}$$

Simplifying (99)

$$S_d(n\gamma) = 0 \tag{100}$$

Since γ and n are assumed to be strictly positive (101)

$$S_d = 0 \tag{102}$$

Thus, this equilibrium is satisfied only under the same conditions as iii (103)

6 Appendix II - Steady State Stability Calculations

The general form of the Jacobian matrix is:

$$\begin{bmatrix} \pi_c - \bar{\pi} + S_c(\frac{\delta(\pi_c - \bar{\pi})}{\delta S_c}) & S_c(\frac{\delta(\pi_c - \bar{\pi})}{\delta S_d}) \\ S_d(\frac{\delta(\pi_d - \bar{\pi})}{\delta S_c}) & \pi_d - \bar{\pi} + S_d(\frac{\delta(\pi_d - \bar{\pi})}{\delta S_d}) \end{bmatrix} \tag{104}$$

For any potential equilibrium, it must be the case that either $S_c = 0$ or $\pi_c - \bar{\pi} = 0$. Similarly, it must be the case that either $S_d = 0$ or $\pi_d - \bar{\pi} = 0$. Thus, we can make the following simplification:

$$\begin{bmatrix} \pi_c - \bar{\pi} + 0 & 0 \\ 0 & \pi_d - \bar{\pi} + 0 \end{bmatrix} \tag{105}$$

Thus, the trace and the determinant for each matrix must be as follows:

$$Tr(J) = J_{11} + J_{22} = \pi_c + \pi_d - 2\bar{\pi} \tag{106}$$

$$Det(J) = J_{11}J_{22} - 0 = (\pi_c - \bar{\pi})(\pi_d - \bar{\pi}) \tag{107}$$

An equilibrium has local asymptotic stability if and only if the Jacobian at that point has a positive determinant and negative trace. We can use this information to calculate the stability conditions for every potential equilibrium.

6.1 The Guard Model

6.1.1 Cooperators Only - $S_d = 0$, $S_e = 0$, $\bar{\pi} = \pi_c$

$$Tr(J) = \pi_d - \pi_c = (x_d - x_c)(A(X) - w)(1 - \alpha) + n\delta(0) \rightarrow Tr(J) > 0 \tag{108}$$

Thus, the equilibrium is always unstable.

6.1.2 Defectors Only - $S_c = 0$, $S_e = 0$, $\bar{\pi} = \pi_d$

$$Tr(J) = \pi_c - \pi_d = (x_c - x_d)(A(X) - w)(1 - \alpha) + n\delta(0) \rightarrow Tr(J) < 0 \tag{109}$$

$$Det(J) = (\pi_c - \bar{\pi})(\pi_d - \pi_d) = 0 \tag{110}$$

The trace is strictly negative, so the equilibrium is always stable.

6.1.3 Enforcers Only - $S_c = 0$, $S_d = 0$, $\bar{\pi} = \pi_e$

$$Tr(J) = \pi_c + \pi_d - 2\pi_e = (x_d + x_c)(A(X) - w)(1 - \alpha) - n\delta - 2(0 - 0) \tag{111}$$

$$Tr(J) < 0 \rightarrow (x_d + x_c)(A(X) - w)(1 - \alpha) < n\delta \tag{112}$$

$$Det(J) = \pi_c\pi_d - 0(0) = \pi_c\pi_d \tag{113}$$

$$\text{Since } \pi_c > 0 \text{ by definition} \tag{114}$$

$$Det(J) > 0 \rightarrow \pi_d > 0 \rightarrow x_d(1 - \alpha)(A(X) - w) > \delta n \rightarrow \pi_d > 0 \tag{115}$$

$$\tag{116}$$

The determinant condition implies the trace condition. The determinant is indefinite, so the equilibrium is conditionally stable.

6.1.4 No Cooperators - $S_c = 0$, $\pi_d = \pi_e = \bar{\pi}$

$$Tr(J) = \pi_c - \pi_d \tag{117}$$

$$Tr(J) < 0 \rightarrow n\delta(1 - S_d) > (x_d - x_c)(1 - \alpha)(A(X) - w) \rightarrow S_d < 1 - \frac{(x_d - x_c)(1 - \alpha)(A(X) - w)}{n\delta} \tag{118}$$

$$Det(J) = (\pi_c - \pi_d + 0)(\pi_d - \pi_d + 0) - 0 = (\pi_c - \pi_d)(0) = 0 \tag{119}$$

$$\tag{120}$$

The trace is indefinite, so the equilibrium is conditionally stable.

6.1.5 No Defectors - $S_d = 0$, $\pi_c = \pi_e = \bar{\pi}$

$$Tr(J) = \pi_d - \pi_c = (x_d - x_c)(A(X) - w)(1\alpha) - n\delta(1 - S_c) \tag{121}$$

$$Tr(J) < 0 \rightarrow S_c < 1 - \frac{(x_d - x_c)(A(X) - w)(1\alpha)}{n\delta} \tag{122}$$

$$Det(J) = (\pi_c - \pi_c + S_c(0))(\pi_d - \pi_c + 0) - 0(0) = 0 \tag{123}$$

The trace is indefinite, so the equilibrium is conditionally stable.

6.1.6 No Enforcers - $S_e = 0$, $\pi_c = \pi_d = \bar{\pi}$

$$Tr(J) = \pi_d - \pi_c = (x_d - x_c)(A(X) - w)(1\alpha) - 0 \tag{124}$$

The trace is strictly positive, so the equilibrium is always unstable.

All Payoffs equal - $\pi_c = \pi_d = \pi_e = \bar{\pi}$

$$Tr(J) = \pi_c + \pi_d - 2\bar{\pi} = 0 \tag{125}$$

$$Det(J) = (\pi_c - \bar{\pi})(\pi_d - \bar{\pi}) = 0 \tag{126}$$

$$\tag{127}$$

Both the trace and determinant are zero, so the equilibrium is always unstable.

6.2 The Vigilante Model

6.2.1 Cooperators Only - $S_d = 0$, $S_e = 0$, $\bar{\pi} = \pi_c$

$$Tr(J) = \pi_d - \pi_c = (x_d - x_c)(A(X) - w)(1 - \alpha) - n\delta(0) \rightarrow Tr(J) > 0 \tag{128}$$

The trace is strictly positive, so the equilibrium is always unstable.

6.2.2 Defectors Only - $S_c = 0$, $S_e = 0$, $\bar{\pi} = \pi_d$

$$Tr(J) = \pi_c - pi_d = (x_c - x_d)(A(X) - w) - n\delta(0) \rightarrow Tr(J) < 0 \tag{129}$$

$$Det(J) = (\pi_c - \bar{\pi})(\pi_d - \pi_d) = 0 \tag{130}$$

The trace is strictly negative, so the equilibrium is always stable.

6.2.3 Enforcers Only - $S_c = 0$, $S_d = 0$, $\bar{\pi} = \pi_e$

$$Tr(J) = \pi_d - \pi_c = (x_d - x_c)(A(X) - w) - n\delta \tag{131}$$

$$Tr(J) < 0 \rightarrow (x_d - x_c)(A(X) - w) < n\delta \tag{132}$$

$$Det(J) = (\pi_c - \pi_e)(\pi_d - \pi_e) = n\delta S_d(\pi_d - \pi_e) = 0 \tag{133}$$

$$\tag{134}$$

The trace is indefinite, so the equilibrium is conditionally stable.

6.2.4 No Cooperators - $S_c = 0$, $\pi_d = \pi_e = \bar{\pi}$

$$Tr(J) = \pi_c - \pi_e = n\delta S_d > 0 \tag{135}$$

$$Det(J) = (\pi_c - \pi_d)(0) = 0 \tag{136}$$

$$\tag{137}$$

The trace is strictly positive, so the equilibrium is unstable unless there are also no defectors (considered separately).

6.2.5 No Defectors - $S_d = 0$, $\pi_c = \pi_e = \bar{\pi}$

$$Tr(J) = \pi_d - \pi_c = (x_d - x_c)(A(X) - w) - n\delta(1 - S_c) \tag{138}$$

$$Tr(J) < 0 \rightarrow S_c < 1 - \frac{(x_d - x_c)(A(X) - w)}{n\delta} \tag{139}$$

$$Det(J) = (\pi_c - \pi_c + S_c(0))(\pi_d - \pi_c + 0) - 0(0) = 0 \tag{140}$$

The trace is indefinite, so the equilibrium is conditionally stable.

6.2.6 No Enforcers - $S_e = 0$, $\pi_c = \pi_d = \bar{\pi}$

$$xc(A(X) - w) < xd(A(X) - w) \rightarrow \pi_c < \pi_d | S_e = 0 \tag{141}$$

$$\tag{142}$$

Since cooperator and defector payoffs cannot be equal without enforcers, no such equilibrium exists.

6.2.7 All Payoffs equal - $\pi_c = \pi_d = \pi_e = \bar{\pi}$

$$Tr(J) = \pi_c + \pi_d - 2\bar{\pi} = 0 \tag{143}$$

$$Det(J) = (\pi_c - \bar{\pi})(\pi_d - \bar{\pi}) = 0 \tag{144}$$

$$\tag{145}$$

Both the trace and determinant are zero, so the equilibrium is always unstable.