A Theory of Cooperation with Self-Commitment Institution

Francesco Lancia and Alessia Russo[1]

August 2^{nd}, 2013

Abstract

Existing theories of cooperation with repeated interactions under imperfect monitoring cannot adequately explain observed trust in the presence of weak enforcement institutions. We introduce the concept of Self-Commitment Institution (SCI) and embed it in an otherwise standard framework to highlight a novel mechanism. When SCI is enforced, agents voluntarily undertake a perfectly observable and costly action, which mitigates agents' inclinations toward opportunistic behavior. We characterize social norms both with and without SCI and provide necessary and sufficient conditions under which the former are socially desirable. We find that (i) when agents are sufficiently patient, players who conform to social norms with SCI are more willing to cooperate, even after realization of a negative event; (ii) intergenerational cooperation is easier to sustain with (without) SCI in the presence of weak (strong) monitoring institutions; (iii) for any level of monitoring, the larger the value of future exchanges, the lower is the value of SCI. In addition, we investigate the role of memory and the impact of growth, yielding the following additional findings: (iv) bounded memory does not preclude the enforceability of trust under SCI; (v) productive SCI sustains Pareto superior equilibria for any quality of external enforcement. Our results are broadly consistent with patterns of cooperation and observable practices in ongoing organizations.

Keywords: Cooperation, Enforcement, Imperfect Public Monitoring, Self-Commitment Institution, Social Norms.

JEL Classification: C72, C73, H40, O43, Z1.

[1]Francesco Lancia, University of Vienna, Email: francesco.lancia@univie.ac.at. Alessia Russo, University of Oslo, Email: alessia.russo@econ.uio.no. We are grateful for valuable comments from Graziella Bertocchi, Michele Boldrin, Bård Harstad, David K. Levine, Anirban Mitra, Nicola Pavoni, Nicola Persico, Karl Schlag, Paolo Siconolfi, Andreas Uthemann and Timothy Worrall. We also thank participants at the 2nd Personnel Economics and Public Finance conference in Ravello, the 11th SAET meeting in Faro, and the 17th ISNIE conference in Firenze, as well as the seminar participants at the Universities of Bergen, Bologna, Firenze, Johannesburg, Napoli, Oslo, and Vienna for useful discussions. We acknowledge the EIEF research grant for financial support. All errors are our own. This paper has circulated in an earlier version under the title "Self-Commitment Institution and Cooperation in Overlapping Generations Games".

1 Introduction

The study of the challenges involved in the sustenance of mutual cooperation among self-interested agents has long occupied a conspicuous position in economics. It is clear that in some organizations, members cooperate more willingly and refrain from free riding than in others. The existence and effectiveness of external enforcement institutions – like courts, legal rules, and regulators – are crucial parts of the explanation but not the sole causes. For elucidation of this point, consider the following Figure.

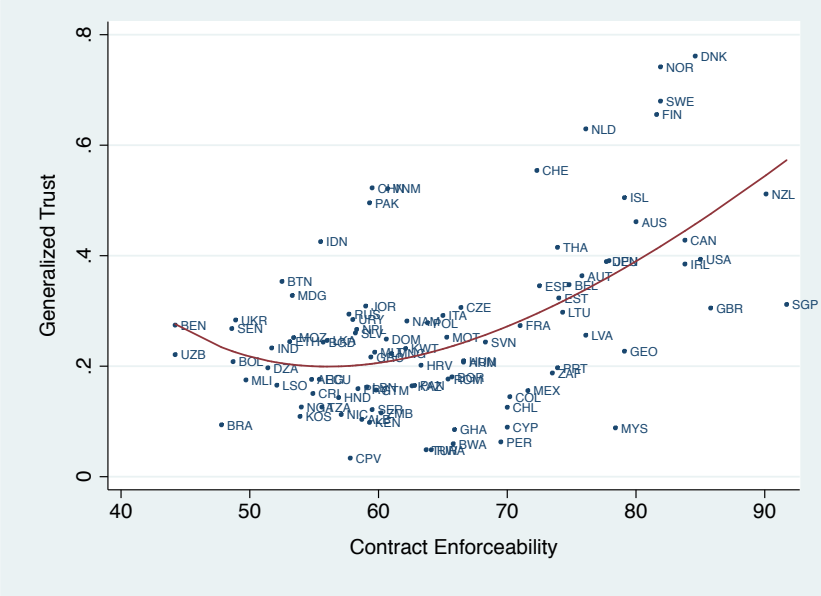


Figure 1: Trust and Enforcement

Figure 1 presents – for a cross section of 100 countries – a plot of *generalized trust* against *contract enforceability*.[2] Not surprisingly, we observe that better contract enforceability is associated with a higher degree of generalized trust. Intuitively, a more effective legal system, through easier detection of deviations, discourages agents from behaving opportunistically; this, in turn, fosters mutual trust. As a mirror-image type of argument, we should expect lower trust in countries with weaker contract enforcement. However, the data suggest an inverse hump-shaped relationship between the two variables. Interestingly, some countries that face low-powered contractual incentives experience relatively high trust. For example, Benin and France differ significantly in terms of contract enforceability, but the level of generalized trust is similar in the two countries.

This Figure raises two fundamental questions. First, why are there such large differences in trust

[2]The variable "Contract Enforceability" in 2009 provides a measure of the efficiency of the judicial system and is obtained from the Doing Business Project by the World Bank, http://www.doingbusiness.org/data. The variable "Generalized Trust" for the wave 2005-2009 is drawn from the World Value survey. The question runs as follows: *"Generally speaking, would you say that most people can be trusted, or that you can't be too careful when dealing with others?"* The trust variable takes a value of 1 if the respondent answers *"Most people can be trusted"* and 0 if he/she answers "*Need to be very careful*".

across societies? Second, could there be informal institutions that individuals rely upon – especially under fragile legal structures – in dealing with potential situations of mistrust? In this study, we investigate the existence of such an alternative institution, which we call Self-Commitment Institution, and present a theory of cooperation that can explain the aforementioned empirical relationships.

In most real-life circumstances, the potential for collective action is confronted with problems of trust and commitment, which are particularly severe when agents do not observe – or equivalently cannot verify – each other's individual contributions. In these circumstances, agents face strong temptations to act against the public interest. Only when individuals can guarantee their participation in a cooperative pursuit is intra-group cooperation more likely to emerge. However, as no court can effectively pre-commit parties to cooperative behavior, individual commitment must be both *self-enforcing* and *credibly perceived* by other community's members.

Our main idea is driven by the following simple observation. Suppose that, in addition to hidden actions, agents can take observable and verifiable actions, due to legal customs or social conventions.[3] These actions could either partially convey hidden information, directly modifying the probability distribution of outcomes, or could have a more subtle effect: they could, *at the margin*, affect incentives to take imperfectly observable cooperative actions. In the latter case, if agents voluntarily undertake such costly and observable actions, then they endogenously undermine their short-term gains from opportunistic behavior and necessarily rely more heavily on the long-term benefits of cooperation. As a result, *the observable action ensures an endogenous and credible commitment to cooperation over unobservables*. Indeed, given the public nature of the information, all community members correctly internalize such changes in the individual incentive structure and properly adjust their retaliation responses. Ultimately, the mechanism leads to higher mutual trust and stronger long-term interpersonal cooperation.

In this paper, we pursue this idea formally by analyzing a standard model of cooperation with repeated interactions and imperfect public monitoring along the lines of Mailath and Samuelson (2006). As a main departure, we allow agents to coordinate their play on the basis of a perfectly observable noncooperative action in addition to the public signals generated by players' past cooperative behavior. The final aim is to characterize the range of situations in which adoption of social norms prescribing the activation of observable noncooperative actions is socially desirable.

Our special focus is on cooperation within ongoing organizations. Therefore, we frame the model as an ongoing economy with an overlapping generation structure inhabited by homogeneous and selfish agents. Individuals live two periods: young and old.[4] The payoff for each generation is de-

[3]Prominent examples are: private education and child-care within the family; blood donation, philanthropy, and civil service or voluntary work in the community; engagement in rituals or adherence to prohibitions in religious organizations; eco- and social-friendly business practices in firms; and so on.

[4]Various models of infinitely repeated interactions among immortal agents have been used previously for the study of cooperation in organizations. However, members of long-lived institutions – like firms, churches, governments, or trade unions - change over time and, after having entered the organization, behave according to preexisting rules. The seminal papers by Cremer (1986) and Kreps (1990) provide a comprehensive framework with overlapping generations for thinking about this issue. These authors build their analysis on the assumption that formal contracts are costly or defective in many situations. Such costs are associated with the costs of monitoring and enforcing the agreement. By contrast, imperfect monitoring is fundamental in the construction of our model. We thus explore new implications of cooperation in organizations.

termined by its interaction with the next-period generation. Each young agents decides whether to cooperate with an elderly counterpart, subject to self-enforcement requirements that the expected gains from cooperation outweigh the one-time gain from defection. As a starting point, consider the case where agents comply with social norms that do not require the activation of observable actions. Thus, future generations only coordinate their play on the basis of a noisy signal of previous players' cooperative behavior. In this scenario, players implement reward/punishment schemes, which are necessarily inefficient. With some probability, they punish non-defecting players and cooperate with guilty ones. For the purpose of characterization, let now introduce the possibility of an alternative organization, whose members comply with social norms that require coordination of individual play *also* with respect to a perfectly observable and costly activity undertaken by previous generations. We call this activity a *self-commitment action* when it satisfies the following decreasing difference property: when taking self-commitment actions, the young reduce their gain from ending cooperation with the old and, thereby credibly signal their intention to cooperate. Accordingly, the institution that fosters this activity is labeled a Self-Commitment Institution (hereafter SCI). One can think of observable actions as purely wasteful activities, which are enforced by a simple reputational mechanism that prescribes reversion to generational autarky in case of defection. In most cases, the possibility of self-commitment actions can be thought of as free disposal of endowment (in terms of time or wealth) or as any costly signal.

A self-commitment action has a two-fold impact on individual welfare. On the one hand, its costly and wasteful nature diminishes the overall gain from intergenerational exchange. On the other hand, its marginal impact on individual cooperative incentives fosters mutual trust. Such a double-edged sword implies that the social desirability of SCI ultimately depends on the fundamentals of the model. A key parameter of the analysis is the likelihood ratio, which measures the effectiveness of the monitoring technology. We show that, if the likelihood ratio is below a certain threshold, then the best self-enforcing governance mode is SCI. On the other hand, if the likelihood ratio is sufficiently high, then social norms without SCI are socially desirable. The likelihood ratio can be interpreted as a proxy for the quality of institutions responsible for external enforcement. Hence, our model predicts that SCI is more likely to be adopted by organizations with low levels of explicit contractual incentives. It also permits organizations to enforce trust when conventional reputational mechanisms would fail. Another relevant parameter of the model is the individual discount factor. We find that, for any given level of monitoring, the more patient agents are, the lower is the social desirability of SCI. Indeed, the larger perceived benefits of future exchanges suffice to provide the adequate incentives to cooperate as well as to abstain from informal institutions of community enforcement such as SCI.

Based on these findings, our model provides a novel theory to explain the above noted differences in trust across organizations. Indeed, in our setting, two societies characterized by different legal structures may exhibit the same patterns of cooperation, provided they rely on different equilibrium regimes: in the presence of weak (strong) monitoring technology, community members comply with social norms with (without) SCI.

Two extensions of the benchmark model address whether the results hinge on particular modeling assumptions. We first address the possibility of one-period memory. In the case of social norms without SCI, we show that generational autarky is the unique self-enforcing outcome, even when players are almost fully patient. This striking result arises from the fact that, due to bounded memory, it is not possible to distinguish punishers from deviators, so that cooperation is not enforced in equilibrium. In contrast, in the case of SCI – by exploiting the property of perfect observability of self-commitment actions – a long-run cooperative outcome can be sustained even under finite memory. It is sufficient to enlarge the self-commitment action space to the point where all of the information encoded in past history is recovered. This result reveals the *signal-driven* nature of institutions such as SCI, compared with the *history-driven* nature of social norms in contexts where self-commitment actions are not contemplated. Second, we extend the model to productive SCI, such as, for instance, parental investment in children's education. We show that the finding, early stated by Rangel (2003), of strategic complementarity between costly investment and intergenerational cooperation holds under imperfect monitoring *only* if costly investment displays self-commitment actions' property of decreasing difference. Indeed, suppose these activities – as opposed to SCI – satisfy an increasing difference property, namely, that when making the costly investment, the young increase their gain from deviations on the cooperative dimension. Then, under weak legal institutions, costly investment and cooperative decisions turn out to be strategic substitutes, in accordance with arguments similar to the one stated above.

Our theoretical framework explains several puzzling facts that are difficult to reconcile with conventional models of cooperation, which rely on reputational arguments. Although the applications are manyl, we limit ourselves to two possibly controversial cases: one involving religious organizations and the other concerning corporate organizations. We interpret time-intensive and wasteful activities, like costly rituals in religious organizations, as credible commitment devices to mitigate free riding problems; socially responsible practices in corporate organizations can be seen in the context of our theory as an optimal communal response in highly volatile financial markets, to foster goodwill and trust among shareholders. In this way, our model complements previous theories of Iannaccone (1992) on religious organization and Baron (2001 and 2010) on corporate organizations.

Our paper is certainly not the first attempt to rationalize differences in mutual trust within organizations. We selectively recall the contributions by Greif, Milgrom, and Weingast (1994), Dixit (2003), Tabellini (2008), and Anderlini and Terlizzese (2012), most based on the pioneering work of Kandori (1992a) and Ellison (1994). This literature shows that, within small homogeneous social groups, informal mechanisms of community enforcement can more effectively support cooperation than formal enforcement mechanisms can. However, as the number of individuals within the fold of this mechanism expands, multilateral punishment becomes more and more ineffective. Indeed, interpersonal communication regarding individuals' past histories becomes prohibitively costly, and *self-governance* – the self-enforceability of mutual exchanges – diminishes. Beyond a certain size, only external enforcement institutions can successfully enlarge the scope of cooperation. Interestingly, other authors such as Ghosh and Ray (1996), Kranton (1996), and Ali and Miller (2013), study the

role of endogenous variable stakes in community enforcement when players are heterogenous, and show that enforcement of cooperation through reputation also helps screen out *bad* players and deter *good* players from shirking. In the long run, this mechanism also fosters greater trust under weak legal structures. Our theory of SCI, through an investigation how certain kinds of commonly-observed past behavior affects the strength of community enforcement, suggests a different explanation for the relationship between the value of exchange and external enforcement institutions. We find the underlying mechanism to be a valuable complement of previous analyses and a further step towards a more general theory of cooperation in ongoing organizations.

Closer to our spirit, Acemoglu and Jackson (2012) explore how occasional "prominent" agents, whose actions are publicly observable by all future agents, affect individual expectations and lead to different regimes of cooperation. In their model, agents must unravel an intergenerational coordination problem, and commitment through "prominent" agents is treated as exogenously given. In our model, players face intergenerational free-riding, and commitment decisions are self-enforcing.

Our study also adds to game-theoretic literature seeking to explain cooperation and collusion in repeated games under imperfect monitoring (e.g., Green and Porter, 1984, Radner, 1986, Abreu, Pearce, and Stacchetti, 1990, and Fudenberg, Levine, and Maskin, 1994). While this literature has mainly sought to characterize cooperative solutions under personal enforcement, our focus is on comparing levels of cooperation in different institutional settings. We contribute to the literature that restricts agents to play strongly symmetric public strategies by analyzing how SCI might sustain stronger cooperation when personal enforcement is interdicted.

Lastly, the paper is related to theoretical contributions on repeated games with overlapping generations of players (e.g., Cremer, 1986, Kandori, 1992*b*, and Bhaskar, 1998). This literature, however, does not in general address questions related to the enforceability of cooperation under imperfect monitoring of players' past performances.

The remainder of the paper is organized as follows. Section 2 describes the model environment. Section 3 introduces the equilibrium concept used and categorizes social norms with and without SCI. The main results pertaining to comparisons of welfare among different institutions are presented in section 4. Section 5 addresses policy implications. Section 6 discusses the two main extensions of the benchmark model, while section 7 explores some applications. Lastly, section 8 concludes. All proofs are contained in the Appendix.

2 The Model

Consider an ongoing economy inhabited by a continuum of selfish agents who imperfectly observe past history. Time is discrete, indexed by $t \in \mathbb{N}$, and runs from zero to infinity. Individuals live two periods. Each agent born at date t plays the role young in period t and the role old in period $t+1$. The payoff of each generation is determined by the interaction with the next-period generation. Figure 2 reports the demographic structure of the economy.

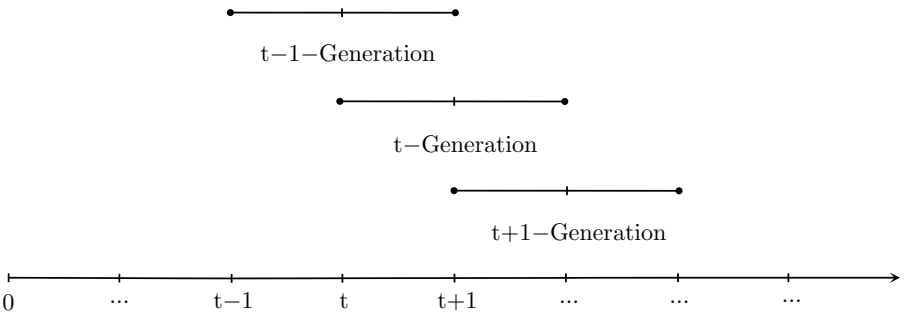


Figure 2: Demographic Structure

2.1 Actions and Payoffs

In the second period of their lives, agents have no future. Hence, when old they adopt their one-shot noncooperative best response. It immediately follows that agents are active only in the first period of their lives. The actions played by the young born in period t are denoted $x_t \in X \equiv \{0, x\}$ and $y_t \in Y \equiv \{0, y\}$, with $x, y > 0$. The ex-post intertemporal utility of an agent depends upon the actions taken at the young age and the actions exerted by the next-period generation. Define the additively separable payoff function, $\upsilon : X \times X \times Y \times Y \to \mathbb{R}$, as follows:

$$\upsilon(\boldsymbol{x}, \boldsymbol{y}) = u(x_t, y_t) + \delta\omega(x_{t+1}, y_{t+1}) \tag{1}$$

where $\delta \in (0, 1)$ is the individual discount factor, $\boldsymbol{x} \equiv (x_t, x_{t+1})$ and $\boldsymbol{y} \equiv (y_t, y_{t+1})$ are the vectors of actions.

Assumption 1 (Preferences) *$u : X \times Y \to \mathbb{R}$ is a non-increasing function in both x_t and y_t, and $\omega : X \times Y \to \mathbb{R}$ is an increasing function in x_t and a non-increasing function in y_t, $\forall t$.*

According to Assumption 1, the action x_t generates a cost for the current generation and a benefit for the past one (positive intergenerational externalities), whereas the action y_t negatively affects the welfare of both generations. Thus, we identify x_t as a *cooperative action* and y_t as a *costly* and *wasteful* action.[5] When agents are self-sufficient, their reservation payoff corresponds to $\upsilon^r \equiv u(0, 0) + \delta\omega(0, 0)$.

Assumption 2 (Payoff-Ranking Condition) *For each x_t and y_t, the following payoff-ranking condition holds:*

$$\upsilon\left((x_t, 0), (y, y_{t+1})\right) < \upsilon^r < \upsilon\left((x_t, x), (y, y_{t+1})\right) < \upsilon\left((x_t, x), (0, 0)\right)$$

Assumption 2 requires that the cooperative outcome entailing $\boldsymbol{x} \gg (0, 0)$ Pareto dominates the generational autarky. Due to the wasteful nature of the non-cooperative action, the optimal allocation

[5] In section 6.2, we extend the benchmark model to an environment endowed with a growth-enhancing technology, where y_t acts as a productive activity.

implemented by the social planner with full commitment, i.e., $\arg\max_{\boldsymbol{x},\boldsymbol{y}} v(\boldsymbol{x},\boldsymbol{y})$, requires $\boldsymbol{y} = (0,0)$. As in Bhaskar (1998), this class of games embeds the incentive structure of the prisoner's dilemma in which the unique static Nash equilibrium entails that agents choose $\boldsymbol{x} = \boldsymbol{y} = (0,0)$.

Definition 1 (Decreasing Difference) *The function $u(x_t, y_t)$ has decreasing difference in (x_t, y_t) if the following inequality holds:*

$$u(0,y) - u(x,y) < u(0,0) - u(x,0) \tag{2}$$

The inequality (2) has a simple interpretation: the incremental gain to choosing a lower x_t is smaller when y_t is higher. That is, the benefits of cheating, $u(0,y_t) - u(x,y_t)$, is decreasing in y_t. The symmetric property of the decreasing difference also ensures that $u(x_t,0) - u(x_t,y)$ is decreasing in x_t. A function characterized by such a structure is called supermodular (see Topkis, 1998).

Definition 2 (Self-Commitment Action) *y_t is defined as a self-commitment action, when $u(x_t, y_t)$ has decreasing difference in (x_t, y_t).*

According to Definition 2, y_t plays the role of a *self-commitment action* when the young reduce their marginal gains from deviation on the x-dimension by choosing $y_t > 0$. The functional property of supermodularity captures a preference for greater interdependence and reflects the complementarity relationship between the two actions. The possibility of a self-commitment action can be interpreted in many ways. In section 7, we discuss two prominent applications: costly rituals in religious organizations and social responsibility in corporate organizations.

2.2 Monitoring Technology

Agents have incomplete information about past history. They observe a public signal of previous cooperative actions, $z \in Z \equiv \{G, B\}$, where G stands for good and B for bad, and perfectly observe all the sequence of the y-actions exerted by previous generations. The distribution of public signals conditioned on the x-action, $\pi_{x_t} \equiv \Pr[B|x_t]$, is denoted $\pi_0 \equiv \Pr[B|0]$ and $\pi_x \equiv \Pr[B|x]$. Signals do not directly impact agents' payoffs. Alternatively, it could be assumed that payoffs are contingent on a random output generated by individual cooperative actions. However, as elderly agents are passive, and intergenerational communication is precluded (unlike Anderlini, Gerardi, and Lagunoff, 2010), such a modeling structure would obscure the analysis without modifying the main findings.

Assumption 3 (Monotone Likelihood Ratio Property) *Given $\pi_0, \pi_x \in [0,1]$, the monotone likelihood ratio property requires $\mathcal{L} \equiv \frac{\pi_0}{\pi_x} \in [1, \infty)$.*

Under Assumption 3, the probability of receiving a good signal is positively correlated with the intensity of the individual cooperative action: the larger is x_t, the higher is the probability of generating a good signal. The likelihood ratio $\mathcal{L}$ represents a sufficient statistic of the monitoring technology effectiveness. In the extreme case of $\mathcal{L} = 1$, which implies $\Delta\pi \equiv \pi_0 - \pi_x = 0$, agents are precluded from detecting deviations by observing the signals. By contrast, when $\mathcal{L} \to \infty$, i.e., $\Delta\pi \to 1$, agents perfectly discern previous individual behavior.

Definition 3 (OLG Game) *The pair* (υ, π_{x_t}) *is referred to as an overlapping generation game with imperfect public monitoring and is denoted* $\mathcal{G}(\upsilon, \pi_{x_t})$.

Several applications fit into our theoretical framework. The following parametric example, which we carry throughout the analysis, provides a simple representation of the implicit setup described above.

Example I Consider a consumption-loan model in which the youth cohort evaluates consumption according to a separable additive utility function of the CARA type $\upsilon_t = -e^{-\gamma c_t^1} - \delta e^{-\eta c_{t+1}^2}$, where γ, $\eta > 0$ are the cohort-targeted absolute risk aversion coefficients. c_t^1 denotes consumption at time t, when young, and c_{t+1}^2 represents consumption at time $t+1$, when old. Agents are born with w units of wealth endowment and spend y_t units on relation-specific investment with no direct benefits. In the first period of their lives, agents use their net endowment, $w - y_t$, to consume and finance the wealth of the elderly cohort through both a mandatory transfer, θ, and a voluntary one, x_t. When old, agents consume their total income equal to the sum of the transfers their children pass on to them. Therefore, the individual budget constrains for the youth and elderly agents are $c_t^1 = (w - y_t)(1 - \theta - x_t)$ and $c_{t+1}^2 = (w - y_{t+1})(\theta + x_{t+1})$, respectively. It is easy to show that there exists a threshold level $\hat{\gamma} \in \mathbb{R}_+$, such that if $\gamma < \hat{\gamma}$, then y_t acts as self-commitment action, as described in Definition 2. □

3 Information Structure and Equilibrium

We adopt the best sustainable strongly symmetric *Public Perfect Equilibrium* (hereafter PPE) as equilibrium concept of $\mathcal{G}(\upsilon, \pi_{x_t})$. Let μ_t denote an independent and identically distributed public random variable extracted from a uniform distribution over the unit interval. Like a sunspot, the public randomization device is an extrinsic random variable, which is relevant information to the coordination of individual behavior, without relation to fundamentals such as preferences, technology, or endowments. We refer to $\mu^t \equiv (\mu_0, \mu_1, ..., \mu_t)$ and $z^t \equiv (z_1, z_2, ..., z_t)$ as the vectors of public randomization devices and public signals generated by x-actions until time t, respectively. Furthermore, $y^t \equiv (y_0, y_1, ..., y_{t-1})$ is defined as the vector of individual y-actions taken until time t. Hence, the public history observed by t-generation is $h^t \equiv (\mu^t, z^t, y^t) \in \mathcal{H}^t$, where $\mathcal{H}^t$ denotes the set of possible public histories until time $t > 0$ and $\mathcal{H} \equiv \bigcup_{t \geq 0} \mathcal{H}^t$ with $\mathcal{H}^0 \equiv \varnothing$. For each $s \leq t$, we refer to $\mu_s(h^t)$, $z_s(h^t)$, and $y_s(h^t)$ as the realizations of μ_s, z_s, and y_s in the public history h^t, respectively. For each t-generation, the public strategies are the mappings $\theta^x : \mathcal{H}^t \to X$ and $\theta^y : \mathcal{H}^t \to Y$, such that $\theta^x(h^t) \in X$ and $\theta^y(h^t) \in Y$. The strategy profiles $\boldsymbol{\theta}^x \equiv (\theta^x(h^t))_{t=0}^{\infty}$ and $\boldsymbol{\theta}^y \equiv (\theta^y(h^t))_{t=0}^{\infty}$ specify the amount of x_t and y_t for any possible history. As all individuals are ex-ante identical and can only be distinguished through their histories, we restrict the analysis to symmetric strategies – each player follows the same strategy after every history.

Rewriting Eq. (1), the ex-ante payoff for each t-generation, conditional on the history h^t, is:

$$\upsilon\left(x_t, y_t | h^t\right) = u(x_t, y_t) + \delta \mathbb{E}_{\mu}\left[\omega\left(\theta^x\left(h^{t+1}\right), \theta^y\left(h^{t+1}\right)\right) | h^t\right]$$

where $\mathbb{E}_{\mu}[\cdot|h^t]$ is the expectation operator, conditional on information at time t.

Definition 4 (PPE) *$\forall t \geq 0$, a strategy profile $(\boldsymbol{\theta}^x, \boldsymbol{\theta}^y)$ is a PPE of $\mathcal{G}(\upsilon, \pi_{x_t})$ if:*

(**i**) *$\theta^x(h^t)$ and $\theta^y(h^t)$ are public strategies;*

(**ii**) *$\forall h^t$, $\theta^x(h^t)$ and $\theta^y(h^t)$ are Nash equilibria from that date onwards.*

An equilibrium strategy of $\mathcal{G}(\upsilon, \pi_{x_t})$ is usually referred to as a *social norm* or implicit contract. It coordinates agents' expectations by prescribing specific behavioral rules and reduces transaction costs in interactions that possess multiple equilibria. In contrast to legal norms or explicit contracts that are typically enforced by a third party, social norms are self-enforced. Indeed, after they are established, they remain in force because agents conform to those rules, given the expectation that others will also comply with them.

In a framework characterized by strategic interactions, social norms are typically sustained by two types of self-enforcement mechanisms: personal and community enforcement. Under personal enforcement, a cheater will only face retaliation by their victim. By contrast, under community enforcement, all members of the society react to any individual deviation. In our environment, due to the demographic structure of the game, personal enforcement is interdicted. Therefore, social norms are sustained only through community enforcement. Under imperfect monitoring, we investigate the strategic role played by *Self-Commitment Institution* (hereafter SCI) – referred to as the institution prescribing the activation of self-commitment actions – in fostering community enforcement of cooperative behavior. To fully characterize the equilibrium outcome, we partition the set of strategies into two subsets corresponding to different social norms, that are with and without SCI.

Definition 5 (Social Norm without SCI) *A social norm without SCI is a PPE characterized by strategies measurable w.r.t. h^t/y^t:*

$$\left\{\theta^x\left(h^t\right) = \theta^x\left(\tilde{h}^t\right) \text{ and } \theta^y\left(h^t\right) = \theta^y\left(\tilde{h}^t\right) \text{ if } h^t \equiv \left(\mu^t, z^t, y^t\right) \text{ and } \tilde{h}^t \equiv \left(\mu^t, z^t, \tilde{y}^t\right) \ \forall\, y^t \neq \tilde{y}^t\right\}$$

In this scenario, even if agents observed past self-commitment actions, they would condition the continuation play only on the history of the public signals generated by cooperative actions.

Definition 6 (Social Norm with SCI) *A social norm with SCI is a PPE characterized by strategies measurable w.r.t. h^t:*

$$\left\{\exists\, h^t \equiv \left(\mu^t, z^t, y^t\right) \text{ and } \tilde{h}^t \equiv \left(\mu^t, z^t, \tilde{y}^t\right) \text{ with } y^t \neq \tilde{y}^t \mid \theta^x\left(h^t\right) \neq \theta^x\left(\tilde{h}^t\right) \text{ or } \theta^y\left(h^t\right) \neq \theta^y\left(\tilde{h}^t\right)\right\}$$

Unlike Definition 5, when social norms with SCI are enforced, players recognize the histories of both public signals and self-commitment actions as relevant information.

As is widely known (see Mailath and Samuelson, 2006), public strategies have a tractable recursive structure, easily recovered through their automaton representations. Such a structure enables us to encode the entire history of past signals, a history that grows over time into a finite

number of states. Through this formulation, social norms are described by the collection $\Xi \equiv \left\{\Phi, \phi_0, \left(\sigma^i\right)_{i\in\{x,y\}}, Q(\cdot)\right\}$, where Φ is the state space whose elements ϕ_{t+1} are drawn from a distribution $Q(\phi_{t+1}|\phi_t, z_{t+1}, y_t) \in \Delta(\Phi)$, with ϕ_0 drawn from an initial given distribution $q_0 \in \Delta(\Phi)$. Given the current state, ϕ_t, and the current actions, (x_t, y_t), the distribution over the next-period state, ϕ_{t+1}, takes the form $q(\phi_{t+1}|\phi_t, x_t, y_t) = \sum_{z_{t+1}\in Z} Q(\phi_{t+1}|\phi_t, z_{t+1}, y_t) \Pr(z_{t+1}|x_t)$. The t-generation's decision rules, which associate states with action profiles, are $\sigma^x(\phi_t) \in X$ and $\sigma^y(\phi_t) \in Y$. Hence, the decision profiles are $\boldsymbol{\sigma}^x \equiv (\sigma^x(\phi_t))_{t=0}^{\infty}$ and $\boldsymbol{\sigma}^y \equiv (\sigma^y(\phi_t))_{t=0}^{\infty}$. Further, we define $\Gamma_\kappa \subseteq \mathbb{R}$ as the set of equilibrium payoffs with memory κ, i.e., $\Gamma_\kappa \equiv \{\upsilon_{\phi_t} : \Phi \to \mathbb{R}, \forall \phi_t \in \Phi\}$, where $\upsilon_{\phi_t} = u(\sigma^x(\phi_t), \sigma^y(\phi_t)) + \delta \sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1}) q(\phi_{t+1}|\phi_t, \sigma^x(\phi_t), \sigma^y(\phi_t))$.

4 Social Norms under Different Institutional Settings

In this section, we provide the roadmap of our analysis, highlighting the parametric conditions under which social norms with and without SCI are (i) self-enforced, i.e., players are effectively deterred from opportunistic behavior, and (ii) socially desirable.

As is well recognized, repeated interactions generate multiple equilibria. However, equilibrium selection is not an issue in this analysis. We confine our attention to the characterization of the best sustainable equilibrium payoff within each category of social norms. It is defined as the highest equilibrium payoff that individuals can attain when they entertain the most optimistic expectation of the other players' behavior. Its general formula is given by:

cooperative payoff - agency cost

Under imperfect monitoring and a small signal space, as in $\mathcal{G}(\upsilon, \pi_{x_t})$, generations implement reward/punishment schemes that are necessarily inefficient. With some positive probability, they punish guiltless players and reward guilty ones. Therefore, the best sustainable equilibrium payoff is bounded away from the efficient cooperative payoff, independently of the discount factor. Such a distance is determined by the agency cost, a measure of the endogenous gain from the current deviation and the cost of monitoring.

4.1 Social Norms without Self-Commitment Institution

We now characterize the set of equilibrium payoffs when agents comply with social norms without SCI, i.e., $\Gamma_\kappa = [\upsilon_{\min}, \upsilon_{\max}]$. It is straightforward to show that the worst sustainable equilibrium payoff of $\mathcal{G}(\upsilon, \pi_{x_t})$ coincides with the generational autarky, i.e., $\upsilon_{\min} = \upsilon^r$. If players know that no one cooperates, independently of the realization of public signals, then it is individually optimal to not cooperate. Clearly, there can be no worse equilibrium than this, because it holds each player at her reservation value. To describe the best sustainable payoff, let us define the strategy Ξ as:

$$\left(\theta^x\left(h^t\right), \theta^y\left(h^t\right)\right) = \begin{cases} (0,0) & \text{if } \iota_{t+1-j} = 0, \forall\, j = 1, ..., n \text{ with } n \geq 1 \text{ odd} \\ (x,0) & \text{otherwise} \end{cases} \tag{3}$$

where $(\theta^x(\varnothing), \theta^y(\varnothing)) = (x, 0)$. For each date $t \geq 1$, ι_t is a flag-function that keeps track of the past history, so that:

$$\iota_t = \begin{cases} 0 & \text{if} \quad z_t = B, \mu_t \geq \mu, \forall\, y_{t-1} \\ 1 & \text{if} \quad [z_t = B, \mu_t < \mu, \forall\, y_{t-1}] \text{ or } [z_t = G, \forall\, \mu_t, y_{t-1}] \end{cases}$$

As long as $\iota_t = 0$, it signals a potential deviation from the cooperative path. To maximize the reward of the punisher and, in turn, enforce long-run cooperation, the equilibrium strategy entails *non-cooperation* when the flags immediately preceding any play are an odd sequence of zeros. For contradiction, suppose that the prescribed sequence were even. Then punishers might be punished in their turn and cooperation would fail. As in Bhaskar (1998), the existence of cooperative equilibria depends crucially on the observability of the entire history of play. In particular, non-cooperation is the unique equilibrium in pure strategies when agents observe at most the signals generated by the last κ generations with κ finite. Hence, memory must be infinite, i.e., $\kappa = \infty$.

Clearly, strategy (3) is not the uniquely feasible strategy. However, we find this strategy convenient because of two appealing properties. First, it has a simple structure. Second, punishment lasts for one period only and only the generation that fails to produce the prescribed level of z_{t+1} and y_t is subject to punishment. An alternative equilibrium strategy is the standard grim-trigger, which prescribes punishment of agents following any bad signal and a sufficiently high realization of the public randomization device. However, as a drawback, it fails to sustain cooperation in the long run. As a final remark, the public randomization device not only simplifies the analysis of the model but also plays a fundamental role. As in Ellison (1994), it fine-tunes the severity – or, alternatively, the duration – of the punishment. Indeed, the lower the endogenous cut-off level μ – a measure of the equilibrium probability that each t-generation assigns to the decisions of future generations to not punish, conditional on the realization of a bad signal – the harsher is the punishment scheme that deters agents from deviating. We will shortly discuss how μ is determined in equilibrium, but for now we assume it is given.

Proposition 1 *Let* $\delta_\infty \equiv \frac{1}{\Delta\pi} \frac{u(0,0)-u(x,0)}{\omega(x,0)-\omega(0,0)}$. *The equilibrium payoff set enforced by the strategy* (3) *is:*

$$\Gamma_\infty = \begin{cases} [v^r, v_{\max}] & \text{if } \delta \geq \delta_\infty \\ \{v^r\} & \text{otherwise} \end{cases}$$

where

$$v_{\max} \equiv u(x,0) + \delta\omega(x,0) - \mathcal{C}(x,0) \tag{4}$$

with $\mathcal{C}(x,0) \equiv \frac{u(0,0)-u(x,0)}{\mathcal{L}-1}$ *denoting the agency cost.*

When $\delta < \delta_\infty$, the unique equilibrium outcome entails $\boldsymbol{x} = \boldsymbol{y} = (0,0)$, whereas when $\delta \geq \delta_\infty$, the optimal equilibrium is larger than the reservation payoff. Because bad signals occur with positive probability along the equilibrium path, it suffices to bound payoffs away from their efficient levels. Such efficiency loss is measured by $\mathcal{C}(x,0)$ in its components $u(0,0) - u(x,0)$ and $\frac{1}{\mathcal{L}-1}$. That is,

larger benefits from deviation or lower informativeness regarding the quality of performance lead to lower ex-ante efficiency. Eq. (4) represents the upper bound of PPE within the class of social norms without SCI. According to Abreu, Pearce, and Stacchetti (1990), under imperfect public monitoring, if the maximal value of a symmetric pure strategy equilibrium is larger than the worst sustainable payoff, then it is achieved for some μ by the grim-trigger strategy. As the equilibrium payoff yielded by the strategy (3) is equivalent to that achieved by the grim-trigger strategy, then we conclude that there can be no preferable equilibrium.

In view of Proposition 1, let us provide the equivalent two-state automaton representation of the social norm described in (3), where $\phi_t \in \Phi = \{\phi_R, \phi_P\}$ for each $t \geq 0$, with ϕ_R denoting the *reward state* and ϕ_P the *punishment state*. For each y_t, the transition probability prescribes:

$$Q\left(\phi_R|\phi_t, z_{t+1}, y_t\right) = \begin{cases} 1 & if \quad \phi_t = \phi_R,\ z_{t+1} = G, \\ \mu & if \quad \phi_t = \phi_R,\ z_{t+1} = B, \\ 1 & if \quad \phi_t = \phi_P,\ \forall z_{t+1} \end{cases}$$

Consistently with Definition 5, a social norm Ξ without SCI is characterized by a transition function that maps the current state to the next-period one by conditioning the probability *only* on the signals generated by past cooperative actions and disregarding the self-commitment activities. As a result, any sequential equilibrium never entails agents taking self-commitment decisions, i.e., $\sigma^y\left(\phi_R\right) = \sigma^y\left(\phi_P\right) = 0$. Furthermore, players cooperate in the reward state and do not cooperate in the punishment state, i.e., $\sigma^x\left(\phi_R\right) = x$ and $\sigma^x\left(\phi_P\right) = 0$. Figure 3 provides a graphical representation of Ξ.

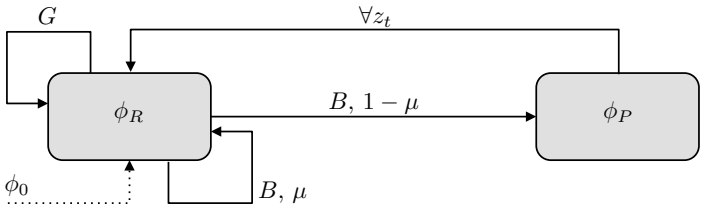


Figure 3: Social Norm without SCI

Agents start to play cooperatively, i.e., $\phi_0 = \phi_R$, and remain in that state until the realization of both a bad signal and a sufficiently high level of μ_s. Next, they play the one-period punishment and, immediately afterward, re-coordinate on the cooperative path, independently of the signal. As ϕ_P is activated on the equilibrium path, Ξ induces boundedness away from the strongly symmetric equilibrium outcome of the efficient cooperative payoff. In the reward state, the intertemporal payoff is as follows:

$$\upsilon_{\phi_R} = u\left(x, 0\right) + \delta \sum_{\phi_{t+1} \in \Phi} \omega\left(\phi_{t+1}\right) q\left(\phi_{t+1}|\phi_R, x, 0\right) \tag{5}$$

where $\sum_{\phi_{t+1} \in \Phi} \omega\left(\phi_{t+1}\right) q\left(\phi_{t+1}|\phi_R, x, 0\right) \equiv \omega\left(x, 0\right) - \pi_x\left(1 - \mu\right)\left[\omega\left(x, 0\right) - \omega\left(0, 0\right)\right]$ is the expected con-

tinuation value in the case of cooperation. In the punishment state, the intertemporal value is:

$$\upsilon_{\phi_P} = u(0,0) + \delta\omega(x,0)$$

A social norm is enforceable – that is, it constitutes a PPE – if and only if each player has an incentive to behave according to the rule. Specifically, it must be true that $(x,0) \succeq_{\phi_R} (0,0)$, i.e., in the reward phase, players must prefer to play $\sigma^x(\phi_R) = x$ rather than deviate to $\sigma^x(\phi_R) = 0$. This implies:

$$\upsilon_{\phi_R} \geq u(0,0) + \delta \sum_{\phi_{t+1} \in \Phi} \omega(\phi_{t+1}) q(\phi_{t+1}|\phi_R, 0, 0) \tag{6}$$

where $\sum_{\phi_{t+1} \in \Phi} \omega(\phi_{t+1}) q(\phi_{t+1}|\phi_R, 0, 0) \equiv \omega(x,0) - \pi_0(1-\mu)[\omega(x,0) - \omega(0,0)]$ is the expected continuation value in the case of defection. Furthermore, it must be true that $(0,0) \succeq_{\phi_P} (x,0)$, i.e., in the punishment phase, players must prefer to play $\sigma^x(\phi_P) = 0$ rather than deviate to $\sigma^x(\phi_P) = x$:

$$\upsilon_{\phi_P} \geq u(x,0) + \delta\omega(x,0) \tag{7}$$

Eq. (7) trivially holds with strict inequality, whereas the incentive compatibility constraint, as in Eq. (6), is satisfied under the following condition:[6]

Condition 1 $(x,0) \succeq_{\phi_R} (0,0)$ *if and only if* $\mu \leq \overline{\mu} \equiv 1 - \frac{1}{\Delta\pi} \frac{u(0,0)-u(x,0)}{\delta(\omega(x,0)-\omega(0,0))}$.

$\bar{\mu}$ indicates the optimal randomization cut-off. It is negatively correlated with both the agent's impatience rate and the level of the agency cost and positively correlated with the loss from being cheated, i.e., $\omega(x,0) - \omega(0,0)$. For contradiction, suppose that μ is strictly larger than $\bar{\mu}$. Then the deterrent effect of the punishment scheme would be too weak to effectively reduce the temptation to deviate. In the opposite case, i.e., μ strictly less than $\bar{\mu}$, υ_{ϕ_R} might be increased, as implied by Eq. (5), without violating the equilibrium condition, Eq. (6). Thus, to yield the best sustainable equilibrium payoff, players entertain the most optimistic expectations consistent with incentive compatibility. Plugging $\mu = \overline{\mu}$ into Eq. (5) yields the maximum payoff as in Eq. (4). To conclude the analysis, we must check that the feasibility conditions are satisfied, i.e., $\overline{\mu} \in [0,1]$.

Condition 2 $\overline{\mu} \in [0,1]$ *if and only if* $\pi_x \leq \pi_0 - \Omega$, *with* $\Omega \equiv \frac{u(0,0)-u(x,0)}{\delta(\omega(x,0)-\omega(0,0))}$.

Clearly, $\overline{\mu}$ is strictly less than one and, under Condition 2, non-negative. By contradiction, if the likelihood ratio were excessively small, i.e., $\pi_x > \pi_0 - \Omega$, then the enforceability constraint, Eq. (6), would not be satisfied and, in turn, the unique sequential rational outcome would entail the

[6] In contrast to the reward state, in the punishment state any possible randomization over the public signals cannot serve to improve ex-ante efficiency. Let $\lambda_s(h^t)$ be the realization of a random variable λ_s extracted from a uniform distribution, where λ is an endogenous threshold level. Two possible randomizations are considered: agents remain in the punishment state with probability λ in the cases of realization of (i) bad signals or (ii) good signals. In both circumstances, it is easy to show that $\lambda = 0$. Intuitively, in the first case, if λ were larger than zero, then agents would have no incentives to punish. By deviating, they increase the probability of generating the good signal and returning to the cooperative path. As a consequence, any announced punishment scheme is not credible, and the unique sequential rational outcome necessarily entails the reservation payoff. In the second case, the punisher yields the highest incentives to punish when $\lambda = 0$.

reservation value. Indeed, in the extreme case, under the harshest punishment scheme, i.e., $\mu = 0$, agents would not be sufficiently discouraged from deviating. Condition 2 is more restrictive than the monotone likelihood ratio property. It implies a reduction of the feasible space – in terms of the effectiveness of monitoring technology – where the implementation of social norms without SCI succeeds in enforcing intergenerational trust.

4.2 Social Norms with Self-Commitment Institution

The objective of this section is to characterize the set of equilibrium payoffs when agents recognize self-commitment actions exerted by previous generations as relevant information. We denote the set of equilibrium payoffs as $\Gamma^*_\kappa = [\upsilon^*_{\min}, \upsilon^*_{\max}]$. Replicating the argument in section 4.1, $\upsilon^*_{\min} = \upsilon^r$. To determine the best sustainable payoff, consider the following strategy, Ξ^*:

$$\left(\theta^x\left(h^t\right), \theta^y\left(h^t\right)\right) = \begin{cases} (0,y) & \text{if } \iota_{t+1-j} = 0, \forall\, j = 1, ..., n \text{ with } n \geq 1 \text{ odd} \\ (x,y) & \text{otherwise} \end{cases} \tag{8}$$

where $\left(\theta^x\left(\varnothing\right), \theta^y\left(\varnothing\right)\right) = (x,y)$. For each date $t \geq 1$, the flag-function ι_t is now defined as:

$$\iota_t = \begin{cases} 0 & \text{if} \quad [z_t = B,\, y_{t-1} = y,\, \mu_t \geq \mu] \text{ or } [y_{t-1} = 0, \forall\, z_t, \mu_t] \\ 1 & \text{if} \quad [z_t = B,\, y_{t-1} = y,\, \mu_t < \mu] \text{ or } [z_t = G,\, y_{t-1} = y, \forall\, \mu_t] \end{cases}$$

The strategy (8) is structurally similar to that reported in (3). An important difference, however, is that the implementation of SCI requires the activation of *two types of punishment*, which differ in the nature of players' deviation. Conditional on the activation of self-commitment actions, the first punishment occurs under the realization of both the bad signal and a sufficiently large value of the public randomization device. The second punishment is executed when previous generations refuse to exert self-commitment actions independently of realization of public signals and the randomization device. Like social norms without SCI, the proposed strategy provides adequate incentives to enforce cooperation in the long run and requires full memory, i.e., $\kappa = \infty$.

Proposition 2 *Let $\delta^*_\infty \equiv \frac{u(0,y)-u(x,y)}{\Delta\pi(\omega(x,y)-\omega(0,y))}$. The equilibrium payoff set enforced by the strategy (8) is:*

$$\Gamma^*_\infty = \begin{cases} [\upsilon^r, \tilde{\upsilon}^*_{\max}] & \textit{if } \delta \geq \delta^*_\infty \\ \{\upsilon^r\} & \textit{otherwise} \end{cases}$$

where

$$\upsilon^*_{\max} \equiv u\left(x,y\right) + \delta\omega\left(x,y\right) - \mathcal{C}^*\left(x,y\right) \tag{9}$$

with $\mathcal{C}^*\left(x,y\right) \equiv \frac{u(0,y)-u(x,y)}{\mathcal{L}-1}$ denoting the agency cost.

For arguments similar to the ones articulated in section 4.1, Eq. (9) represents the upper bound of PPE within the class of social norms with SCI. The benefits of cheating over the expected cost of being cheated, $\frac{u(0,y_t)-u(x,y_t)}{\omega(x,y_{t+1})-\omega(0,y_{t+1})}$, quantitatively determine the range of discount factors for which the feasible set of equilibrium payoffs contains cooperative outcomes. If it were decreasing in $\boldsymbol{y}$, then

there would exist a range of discount factors, $\delta^*_\infty < \delta < \delta_\infty$, such that social norms with SCI would succeed in enforcing cooperation, whereas social norms without SCI would fail. Note that decreasing differences in current utility do not guarantee this outcome *per se*. Therefore, under some parametric conditions, the reverse relationship might hold. Furthermore, note that self-commitment actions – by reducing the current gain from deviations – improve efficiency of the agency cost component, i.e., $\mathcal{C}^*(x,y) < \mathcal{C}(x,0)$.

The social norm described in (8) has a three-state automaton representation, where $\phi_t \in \Phi = \{\phi_R, \phi_{P_1}, \phi_{P_2}\}$ for each $t \geq 0$, with ϕ_R as the *reward state*, and ϕ_{P_1} and ϕ_{P_2} as two different *punishment states* generated by the two alternative deviations. The transition probability prescribes:

$$Q^*(\phi_R|\phi_t, y_t, z_{t+1}) = \begin{cases} 1 \quad \text{if} \quad y_t > 0 & \begin{cases} \phi_t = \phi_R,\, z_{t+1} = G \\ \phi_t \in \{\phi_{P_1}, \phi_{P_2}\},\, \forall z_{t+1} \end{cases} \\ \mu \quad \text{if} \quad y_t > 0 & \phi_t = \phi_R,\, z_{t+1} = B \\ 0 \quad \text{if} \quad y_t = 0 & \forall \phi_t, \forall z_{t+1} \end{cases}$$

Consistently with Definition 6, a social norm with SCI such as Ξ^* is characterized by a transition function that maps the current state to the next-period state by conditioning the probability on both the realization of signals and the self-commitment actions. In equilibrium, the decision rules prescribe $(\sigma^x(\phi_R) = x, \sigma^y(\phi_R) = y)$ and $(\sigma^x(\phi_{P_\tau}) = 0, \sigma^y(\phi_{P_\tau}) = y)_{\tau \in \{1,2\}}$.

Agents start to play cooperatively, i.e., $\phi_0 = \phi_R$, and remain in that state until they observe:

1. A bad signal, $z_t = B$, and a sufficiently high level of the public randomization device, $\mu_t \geq \mu$; next, they move to ϕ_{P_1};

2. The previous player's defection in the self-commitment dimension, $y_{t-1} = 0$, for any possible realization of public signals; next, they move to ϕ_{P_2}.

Afterwards, they play a one-period punishment. Figure 4 graphs the automaton representation of Ξ^*.

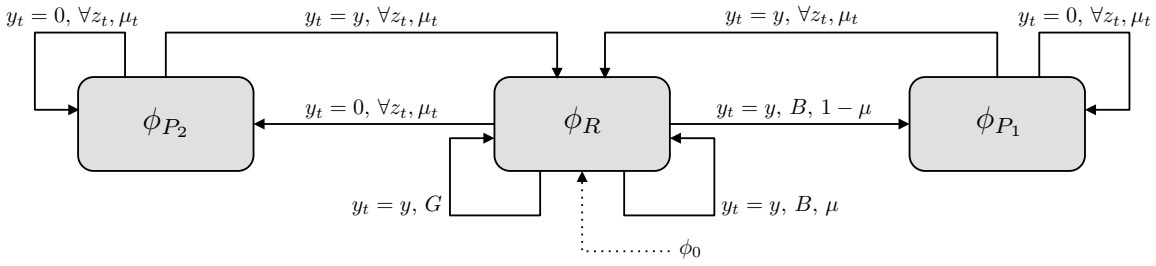


Figure 4: Social Norm with SCI

In the reward state, the intertemporal payoff function is equal to:

$$v_{\phi_R} = u(x,y) + \delta \sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1})\, q^*(\phi_{t+1}|\phi_R, x, y) \tag{10}$$

where $\sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1})\, q^*(\phi_{t+1}|\phi_R, x, y) \equiv \omega(x,y) - \pi_x(1-\mu)[\omega(x,y) - \omega(0,y)]$ is the expected continuation value in the case of cooperation. In the punishment states, the intertemporal values are:

$$v_{\phi_{P_\tau}} = u(0,y) + \delta\omega(x,y)$$

for each $\tau \in \{1,2\}$. The strategy $\tilde{\Xi}^*$ is an equilibrium if and only if in each phase, the incentive compatibility constraints are satisfied. In contrast to $\tilde{\Xi}$, we must check for multi-side deviation incentives in both reward and punishment phases. First, we verify that in the reward phase, $(x,y) \succeq_{\phi_R} (0,y)$, i.e., if agents prefer to play $\sigma^x(\phi_R) = x$ and $\sigma^y(\phi_R) = y$ rather than $\sigma^x(\phi_R) = 0$ and $\sigma^y(\phi_R) = y$:

$$v_{\phi_R} \geq u(0,y) + \delta \sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1})\, q^*(\phi_{t+1}|\phi_R, 0, y) \tag{11}$$

where $\sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1})\, q^*(\phi_{t+1}|\phi_R, 0, y) \equiv \omega(x,y) - \pi_0(1-\mu)[\omega(x,y) - \omega(0,y)]$ is the expected continuation value in the case of defection in the x-action. Furthermore, it must also be true that $(x,y) \succeq_{\phi_R} (0,0)$, i.e., the alternative deviation, $\sigma^x(\phi_R) = 0$ and $\sigma^y(\phi_R) = 0$, is dominated by the cooperative behavior:

$$v_{\phi_R} \geq u(0,0) + \delta \sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1})\, q^*(\phi_{t+1}|\phi_R, 0, 0) \tag{12}$$

where $\sum_{\phi_{t+1}\in\Phi} \omega(\phi_{t+1})\, q^*(\phi_{t+1}|\phi_R, 0, 0) = \omega(0,y)$ is the expected continuation value in the case of defection in both actions. Thus, in contrast to social norms without SCI, the activation of self-commitment actions introduces the possibility of *exiting the game* and, by doing so, affects the realization of the equilibrium outcome. Finally, we check that $(x,y) \succeq_{\phi_R} (x,0)$, i.e., agents must prefer to play $\sigma^x(\phi_R) = x$ and $\sigma^y(\phi_R) = y$ rather than $\sigma^x(\phi_R) = x$ and $\sigma^y(\phi_R) = 0$. Assumption 2 guarantees that $(0,0) \succeq_{\phi_R} (x,0)$ holds. Thus, the latter deviation is disregarded.

Second, we must verify that the players are willing to carry out the punishment, i.e., they have no incentives to deviate from $(\sigma^x(\phi_{P_\tau}) = 0,\, \sigma^y(\phi_{P_\tau}) = y)_{\tau\in\{1,2\}}$. Thus, the following multi-side enforceability constraints must be satisfied:

$$\begin{aligned}
(0,y) \succeq_{P_\tau} (0,0) \quad &\Rightarrow \quad v_{\phi_{P_\tau}} \geq u(0,0) + \delta\omega(0,y) \\
(0,y) \succeq_{P_\tau} (x,0) \quad &\Rightarrow \quad v_{\phi_{P_\tau}} \geq u(x,0) + \delta\omega(0,y) \\
(0,y) \succeq_{P_\tau} (x,y) \quad &\Rightarrow \quad v_{\phi_{P_\tau}} \geq u(x,y) + \delta\omega(x,y)
\end{aligned}$$

for each $\tau \in \{1,2\}$. The first two inequalities hold under the payoff-ranking conditions, while the latter is verified under Assumption 1.

Given the credible punishment threat, in the reward state, agents must always be willing to exert the x- and the y-action for their counterpart; that is, the incentive compatibility requirements given by Eqs. (11) and (12) must hold.

Condition 3 $(x,y) \succeq_{\phi_R} (0,y)$ *if and only if* $\mu \leq \overline{\mu}^* \equiv 1 - \frac{1}{\Delta\pi}\frac{u(0,y)-u(x,y)}{\delta(\omega(x,y)-\omega(0,y))}$.

Condition 4 $(x,y) \succeq_{\phi_R} (0,0)$ *if and only if* $\mu \geq \underline{\mu}^* \equiv \frac{1}{\pi_x}\frac{u(0,0)-u(x,y)}{\delta(\omega(x,y)-\omega(0,y))} - \frac{1-\pi_x}{\pi_x}$.

Clearly, Conditions 3 and 4 must be simultaneously satisfied, i.e., $\mu \in M \equiv \left[\underline{\mu}^*, \overline{\mu}^*\right]$. On the one hand, suppose $\mu > \overline{\mu}^*$. As with social norms without SCI, the weak deterrent power would encourage opportunistic behavior. On the other hand, consider $\mu < \underline{\mu}^*$. Due to excessive punishment, agents would prefer to opt out. Thus, strategy (8) is a PPE only if the feasibility requirements are met, i.e., $\overline{\mu}^* \in [0,1]$ and $M \neq \varnothing$.

Condition 5 $\overline{\mu}^* \in [0,1]$ *if and only if* $\pi_x \leq \pi_0 - \frac{u(0,y)-u(x,y)}{\delta(\omega(x,y)-\omega(0,y))}$.

Condition 6 $M \notin \varnothing$ *if and only if* $\pi_x \leq \Omega^* \pi_0$ *where* $\Omega^* \equiv \frac{\delta(\omega(x,y)-\omega(0,y))-(u(0,0)-u(x,y))}{\delta(\omega(x,y)-\omega(0,y))-(u(0,0)-u(0,y))}$.

For any rate of impatience, Condition 5 holds under Condition 6. Therefore, only the latter will be considered in the following analysis. It drives the main mechanism's insights: the two punishment phases, ϕ_{P_1} and ϕ_{P_2} – although very similar in nature – substantially differ. While ϕ_{P_1} is activated *on the equilibrium* path, as in Ξ, ϕ_{P_2} is as an *out-of-equilibrium* punishment phase. As the y-action is perfectly observable, and players are certainly punished if they decide not to self-commit, in equilibrium, they always choose to sustain that cost. At the same time, the existence of a such an out-of-equilibrium punishment path reduces the intensity of the equilibrium punishment. *When agents coordinate on the y-actions, they reduce their current gain from deviation and, in turn, signal their intention to cooperate, credibly dampening their shirking incentives.* For contradiction, let $M = \varnothing$. Then the punishment state ϕ_{P_2} would be activated along the equilibrium path, and as a result, the best sustainable payoff would collapse at the generational autarky. Finally, computing the maximum payoff of Γ^*_∞ is equivalent to finding the largest μ for which the incentive compatibility constraints and the feasibility constraint are not violated. Plugging $\mu = \overline{\mu}^*$ into Eq. (10) yields the best sustainable payoff as reported in Eq. (9).

4.3 Value of Self-Commitment Institution

What price are agents willing to pay to enforce SCI? In this section, we characterize the conditions under which social norms with SCI outperform social norms without SCI and are thus socially desirable. First, we compare the best sustainable equilibrium payoffs given by Eqs. (4) and (9); second, we determine the equilibrium gain from complying with SCI, i.e., $V(\pi_0,\pi_x) \equiv \upsilon^*_{\max} - \upsilon_{\max}$.

Condition 7 $V(\pi_0,\pi_x) \geq 0$ *if and only if* $\pi_x \geq \Omega^{**}\pi_0$, *with* $\Omega^{**} \equiv \frac{u(x,0)-u(x,y)+\delta(\omega(x,0)-\omega(x,y))}{u(0,0)-u(0,y)+\delta(\omega(x,0)-\omega(x,y))}$.

Self-commitment action has a two-fold impact on individual welfare. On the one hand, its costly nature reduces the cooperative payoff component. On the other hand, its decreasing difference property causes a reduction in agency costs and, in turn, facilitates intergenerational exchange. Such a double-edged sword implies that the social desirability of each social norm ultimately depends on the fundamentals of the model. The latter effect dominates the former when the degree of monitoring is sufficiently ineffective, as stated in Condition 7.

In this section, we question whether weak performance in terms of monitoring justifies the persistence of social norms with SCI, as with costly and wasteful practices that might initially appear inefficient. With this object, we characterize the *necessary* and *sufficient conditions* under which social norms with SCI are optimal. For the sake of exposition, we introduce the following index function.

Definition 7 *The index* $\mathcal{I} : [1, \infty) \to [-1, 1]$ *is defined as* $\mathcal{I}(\mathcal{L}) \equiv \Pr[x|B, \Xi^*, \mathcal{L}] - \Pr[x|B, \Xi, \mathcal{L}]$, *with* $\mathcal{I}(1) = 0$ *and* $\lim_{\mathcal{L}\to\infty} \mathcal{I}(\mathcal{L}) = 0$.

The index $\mathcal{I}(\mathcal{L})$ computes the difference between the optimal enforceable randomization cutoffs of the two social norms. It provides a synthetic measure of the relative ability of SCI to *signal* the intention to cooperate on the x-action: the larger is $\mathcal{I}(\mathcal{L})$, the stronger is the signaling power of SCI. Such an index is strictly positive (negative) when the benefit of cheating over the expected cost of being cheated is decreasing (increasing) in the level of self-commitment actions. When $\mathcal{L} \to \infty$, which implies that $\Pr[x|B, \cdot, \mathcal{L}] = 1$ for each type of social norm, the index is necessarily equal to zero. Remarkably, under some parametric settings, decreasing difference *per se* suffices to make $\mathcal{I}(\mathcal{L})$ strictly positive. This is the case, for example, when $\omega(x_t, 0) - \omega(x_t, y) = 0$, which implies neutrality of the y-action with respect to the elderly payoff, independently of intergenerational exchange.

Proposition 3 (Necessary Condition) *If* $V \geq 0$, *then* $\mathcal{I}(\mathcal{L}) > 0$.
Proof *(See appendix).*

If the index $\mathcal{I}(\mathcal{L})$ were non-positive, then self-commitment actions would fail to signal true intentions to cooperate and, in turn, the value of SCI would necessarily be negative for any degree of monitoring. As in the literature on moral hazard, $\mathcal{L}$ provides a measure of the quality of performance: a lower (higher) likelihood ratio is a poor (rich) informative statistic of the quality of agents' economic activities. For example, in phases of high volatility, i.e., low $\mathcal{L}$, positive performance is not necessarily a reliable indicator of good behavior. In these contexts, agents can face uncertainty by adopting SCI as an *insurance-like* device. SCI can lead to positive evaluations of the performances of other players, who temper their negative judgments and sanctions toward previous generations because of this goodwill. Clearly, if the institution were unable to foster intergenerational trust – and, in turn, endogenously dampen environmental volatility – then players would have no incentives to comply with SCI, as articulated in Proposition 3.

Proposition 4 (Sufficient Condition) *There exists* $\underline{\delta} > 0$ *such that, for any* $\delta \in [\underline{\delta}, 1)$, *the Conditions 2, 6, and 7 are satisfied, which implies that the best symmetric PPE sustained as Social Norm with SCI dominates*

the one sustained as Social Norm without SCI.
Proof *(See appendix)*

According to Proposition 4, when agents are sufficiently patient, social norms with SCI are socially desirable compared to social norms without SCI. It implies that the best equilibrium entails *strategic complementarity* between self-commitment actions and cooperative ones. By contrast, if $\delta < \underline{\delta}$, then the two actions are strategic substitutes.

Let $\Pi \equiv \{(\pi_0, \pi_x) \in [0,1] \times [0,1] \mid \text{Conditions 2, 6, and 7}\}$ denote the parametric space in which both types of social norms enforce cooperation and the value of SCI is positive. Further, define $\bar{\mathcal{L}} \equiv \frac{\bar{\pi}_0}{\bar{\pi}_x}$ as the likelihood ratio corresponding to the highest feasible value of SCI, $\overline{V} \equiv V\left(\bar{\mathcal{L}}\right)$. Then the following result holds.

Corollary 1 (Value of SCI vs Monitoring Technology) *For each $(\pi_0, \pi_x) \in \Pi$, the inequality $\mathcal{L} \geq \bar{\mathcal{L}} = \frac{1}{1-\Omega}$ holds.*
Proof *(See appendix)*

The equilibrium trade-off in Corollary 1 provides a simple theory of the scope of cooperation. In the parametric space Π, the higher is the level of monitoring, the lower is the value of SCI. This result drives a straightforward and testable implication of the model: *intergenerational cooperation is more easily sustained in the presence of* (i) *weak (strong) monitoring institutions and* (ii) *players complying with social norms with (without) SCI*. In section 7, we present ample evidence consistent with our result.

Corollary 2 (Value of SCI vs Value of Exchanges) *For any level of monitoring, the larger is the value of future exchange, the lower is the value of SCI.*

Corollary 2 arises directly from inspection of Condition 7: for a given level of monitoring, the coefficient Ω^{**} is an increasing function of the individual discount factor δ, which leads to a lower value of SCI. Intuitively, the lower the impatience rate, the larger the net present value of future exchange, which provides the appropriate incentives to cooperate as well as to abstain from informal institutions of community enforcement such as SCI.[7] Among others, Greif, Milgrom, and Weingast (1994), and Kranton (1996) have long investigated the relationship between the value of exchange and self-sustaining systems of reciprocal cooperation. Although these studies take a different perspective, their analyses complement ours. Informal institutions that reinforce mutual exchange are interpreted as means for individuals to economize on market search costs. When the size of trade as well as the number of individuals to which this mechanism applies expands, reciprocal-exchange arrangements have fewer benefits, especially when implementation of the mechanism is costly. In this scenario, the authors argue, formal institutions (markets) are an attractive alternative, and reciprocity cannot be enforced.

Figure 5 illustrates the properties of the PPE of $\mathcal{G}\left(\upsilon, \pi_{x_t}\right)$, as discussed above. SCI has a twofold effect on the equilibrium outcome: (i) it expands the parametric space in which intergenerational

[7] The tight link between the value of informal institutions and the value of future exchanges can be further explored by extending SCI to the case of productive investment. We return to this point at the end of section 6.2.

cooperation is sustained, and (ii) it reduces the maximum sustainable payoff yield associated with better monitoring technology. The shaded area of Figure 5 represents the parametric space Π. In Panel (a), we plot the maximum value of SCI as the projection of $V(\pi_0, \pi_x)$ corresponding to the upper envelop of Π over π_0. At $(\bar{\pi}_0, \bar{\pi}_x) = (1, 1-\Omega)$, SCI yields the highest feasible value. As in Corollary 1, the corresponding likelihood ratio – measured by the slope of the ray that crosses the point $(\bar{\pi}_0, \bar{\pi}_x)$ – attains its minimum value. In Panel (b), we provide an alternative geometrical illustration of the equilibrium outcome. Specifically, we plot the value of SCI as a function of $\mathcal{L}$. When $\mathcal{L}$ is below than the minimum likelihood that satisfies Condition 2, i.e., $\mathcal{L} < \frac{1}{1-\Omega}$, then the unique equilibrium outcome sustained by social norms without SCI is necessarily the reservation payoff. Similarly, when $\mathcal{L}$ is below than the minimum likelihood that satisfies Condition 6, i.e., $\mathcal{L} < \frac{1}{\Omega^*}$, then social norms with SCI enforce the reservation value as the unique equilibrium outcome. Therefore, in the interval $\mathcal{L} \in \left(\frac{1}{\Omega^*}, \frac{1}{1-\Omega}\right)$, SCI has a positive value that increases monotonically in the likelihood ratio. This positive relationship reverts for $\mathcal{L} \in \left(\frac{1}{1-\Omega}, \frac{1}{\Omega^{**}}\right)$. However, the value of SCI remains positive. Finally, when $\mathcal{L} > \frac{1}{\Omega^{**}}$, Condition 7 is violated and $V(\pi_0, \pi_x)$ becomes negative.

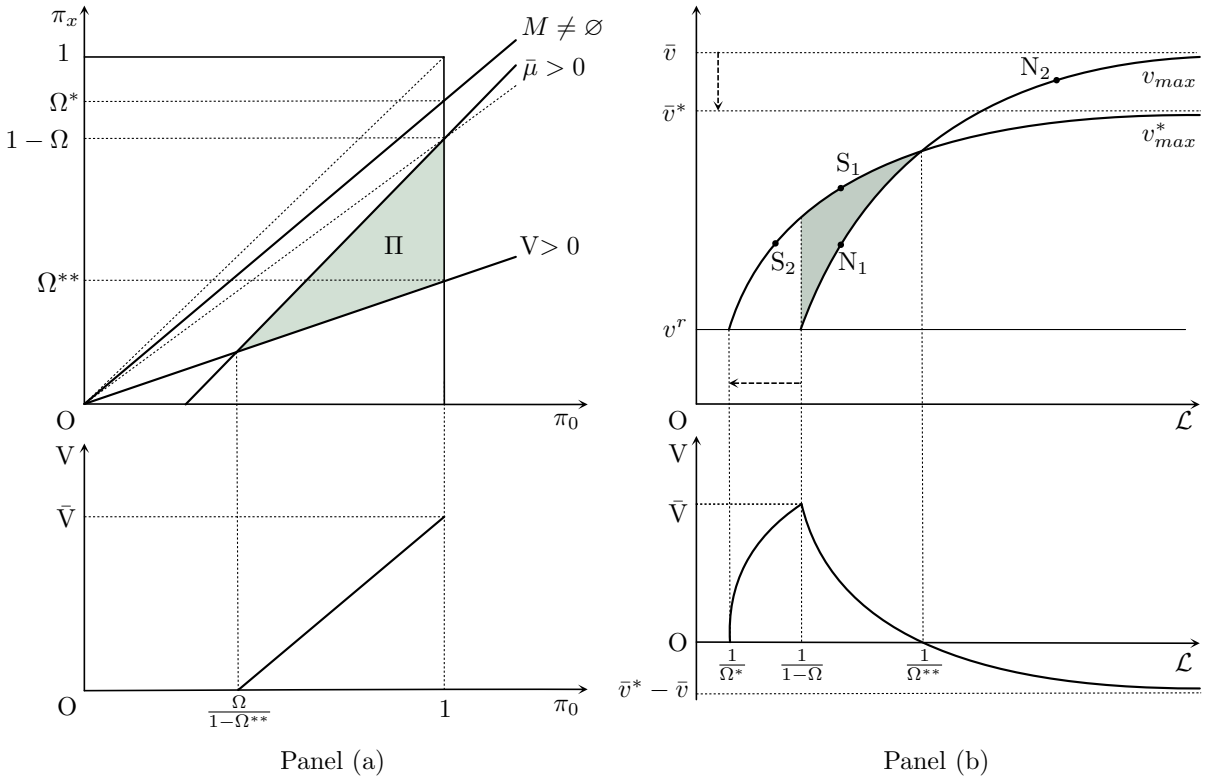


Figure 5: Value of SCI

5 Policy Implications

In recent decades, a massive empirical literature has documented large differences in social trust across societies.[8] Our model provides a novel theory to replicate such differences, distinguishing two main cases. First, two societies characterized by the same parameter settings, but reliant on *differences in conventions*, generate distinct equilibrium regimes. Second, two societies characterized by different parameter structures experience either a given cooperation pattern associated with distinct equilibrium regimes or belong to the same equilibrium regime but have different cooperation levels. We view the latter case as characterized by *differences in fundamentals*.

In our theoretical environment, the structure of equilibrium depends on various parameters. To sharpen the intuition, we single out the role played by the degree of monitoring. From this perspective, our theory calls for distinct types of government interventions aimed at improving trust and efficiency. We consider differences in conventions and differences in fundamentals separately, relying for our discussion on Figure 5.

To begin, consider two societies, denoted N_1 and S_1, endowed with identical monitoring technologies, i.e., $\mathcal{L}^{N_1} = \mathcal{L}^{S_1}$, but differing in their conventions. In N_1, agents comply with social norms without SCI, whereas S_1 implements social norms with SCI. If the level of self-enforced trust in S_1 were higher than that sustained in N_1, then an effective policy intervention would require the government in N_1 to re-focus players' expectations by convincing them of the payoff-relevance of self-commitment actions. For example, policies of this nature might include indoctrination or the implementation of educational programs based on civic culture.

Now consider a scenario with two societies, denoted N_2 and S_2, endowed with different monitoring technologies, specifically, $\mathcal{L}^{N_2} > \mathcal{L}^{S_2}$. Similarly to the previous case, N_2 implements social norms without SCI, whereas S_2 sustains social norms with SCI. As illustrated in Figure 5, due to better technology, N_2 experiences stronger intergenerational trust and higher efficiency than S_2. However, any policy aiming to rectify the low level of cooperation in S_2 by deterring self-commitment decisions would undermine the efficiency in that society. Policies of this nature might include abolishment of SCI or wealth subsidies to compensate for the marginal costs incurred by self-commitment actions. Indeed, in an environment with weak monitoring, these public interventions are inherently wasteful, as they erode the SCI strategic valence by either undermining its effectiveness in signaling intentions to cooperate or inducing even more intensive self-commitment activities, thereby dissipating the initial subsidy.

As discussed above, the parameter $\mathcal{L}$ is a crucial element of the overall analysis. As a measure of monitoring technology effectiveness, it can also be interpreted as an indicator of the quality of institutions responsible for external enforcement. In the introductory section, we stressed the link between the level of generalized trust and the effectiveness of contract enforceability. The above equilibrium provides an analytical foundation for this relationship. From this perspective, our the-

[8] For example, Banfield (1958) and Putnam (1993) provide a prominent example of multiple self-enforcing patterns of cooperation between the south and north of Italy, patterns that they trace to culture and history.

ory conveniently resolves a major debate in the previous literature over whether formal and informal institutions are complements or substitutes.[9] On the one hand, the theory of SCI predicts that better external enforcement - that is, larger $\mathcal{L}$ - corresponds with smaller benefits of cheating and a smaller losses from being cheated and, in turn, an enlarged scope for cooperation for each type of social norm. Hence, formal and informal institutions act as strategic complements, with *trust crowded in by better enforcement*. On the other hand, the model also predicts that societies characterized by weaker external enforcement - that is, lower $\mathcal{L}$ - expand their cooperation possibilities if they implement conventions based on self-commitment actions. As a consequence, *better external enforcement by crowding out values* are substitute for specific types of informal institutions such as SCI. This result can explain why norms of costly and observable activities are often found in societies with histories of weak legal institutions. Historical examples that confirm our claim abound. Section 7 presents some of this evidence along with an exhaustive discussion of the theoretical predictions described above.

Example I Tables 1-3 report comparative statics for γ, w, and θ. For illustrative purposes, fix $x = 9/20$ and $y = 1/10$.

Table 1: Value of Unproductive SCI, w=1, θ=0, η=10 and δ=0.98

γ	$\hat{\gamma}$	$\mathcal{I}$	$\underline{\delta}$	$\overline{\mathcal{L}}$	$V(\overline{\mathcal{L}})$	$\frac{V(\overline{\mathcal{L}})}{v_{\max}}$
0.5	1.400	0.009	0.538	1.188	0.026	0.016
0.6	1.400	0.009	0.669	1.213	0.016	0.010
0.7	1.400	0.008	0.824	1.234	0.007	0.005

Table 2: Value of Unproductive SCI, θ=0, γ=0.6, η=10 and δ=0.98

w	$\hat{\gamma}$	$\mathcal{I}$	$\underline{\delta}$	$\overline{\mathcal{L}}$	$V(\overline{\mathcal{L}})$	$\frac{V(\overline{\mathcal{L}})}{v_{\max}}$
0.8	1.774	0.011	0.580	1.186	0.026	0.016
1	1.400	0.009	0.669	1.213	0.016	0.010
1.2	1.156	0.007	0.800	1.236	0.007	0.005

Table 3: Value of Unproductive SCI, w=1, γ=0.6, η=10 and δ=0.98

θ	$\hat{\gamma}$	$\mathcal{I}$	$\underline{\delta}$	$\overline{\mathcal{L}}$	$V(\overline{\mathcal{L}})$	$\frac{V(\overline{\mathcal{L}})}{v_{\max}}$
0	1.400	0.009	0.669	1.213	0.016	0.010
0.01	1.419	0.012	0.610	1.242	0.012	0.008
0.02	1.439	0.015	0.573	1.277	0.008	0.006

Table 1 indicates that the larger is the intergenerational risk-aversion, i.e., the higher is the ratio $\frac{\eta}{\gamma}$, the stronger is the signaling power of SCI, i.e., the higher is $\mathcal{I}$.[10] The result is intuitive. A larger degree of relative risk-aversion of the elderly cohort implies that the expected loss from being cheated is higher in a society that enforces SCI. This effect entails reduced benefits of cheating and that the

[9] Kranton (1996) has theoretically explored how better external enforcement weakens reputational incentives and, in turn, harms informal institutions. In contrast, Tabellini (2010) has empirically documented the existence of strong positive correlations between the quality of legal institutions and indicators of trust.

[10] The index $\mathcal{I}$ in Tables 1-3 is calculated after normalizing by $\delta\Delta\pi$.

y-action functions unambiguously as a costly signal of self-commitment. Moreover, policies prescribing wealth subsidies – such as an increase in w, as in Table 2 – erode SCI effectiveness in signaling intentions to cooperate – i.e., lower $\mathcal{I}$ – and, in turn, dissipate welfare gains. For example, an endowment subsidy of 0.2 induces a drop of 50% in the maximum efficiency premium yield of SCI, as measured by the ratio $\frac{V(\overline{\mathcal{L}})}{v_{\max}}$. Finally, policies prescribing mandatory transfers, θ, positively affect the signaling power of SCI but crowd out voluntary intergenerational exchanges. As a result, the cost incurred by self-commitment activities turns out to be large compared to the benefits of stronger cooperation. In Table 3, the described effect is captured by the negative relationship between $\mathcal{I}$ and $V\left(\overline{\mathcal{L}}\right)$. □

6 Discussion of Modeling Assumptions

In this section, we discuss the role of modeling assumptions in the results obtained thus far. Specifically, we allow (i) the players to recall one period of previous history and (ii) SCI to be productive.

6.1 Finite Memory

In an ongoing economy inhabited by finite-lived players, the assumption of perfect recall of past history seems especially unreasonable. Indeed, each generation directly observes only the current outcome, with no *direct* memory of past events. Nevertheless, a given individual might form beliefs about past events relying either on *history's footprint*, in the form of a sequence of public signals about past history, or messages from the preceding generation. In our model, although we recognize its theoretical relevance, we have intentionally ruled out the second channel and have focused on the first.[11] Previous sections have shown how a degree of long-run trust is enforced by the two types of social norms – although to differing extents – when each generation fully disposes of history's footprint. In this section, we examine whether memory matters in creating and perpetuating cooperative behavior and whether it differentially affects performances of social norms with and without SCI. For this purpose, we consider the extreme scenario in which agents recall only last-period information. Hence, we study the extent to which efficient equilibrium payoffs can be obtained through one-period memory public strategies.[12]

Let us begin with a simple observation. In our model, due to the dichotomous nature of action and signal spaces, it is generally difficult to determine whether a player has individually deviated. Indeed, deviators and punishers behave similarly by not cooperating with elderly counterparts. However, provision of the right incentives in punishing deviators is crucial to enforcing cooperation: any player is effectively deterred from opportunistic behavior only if he anticipates that the punisher will

[11] Anderlini, Gherardi, and Lagunoff (2010) have investigated how social memory, i.e., the intergenerational transmission of past history, can be systematically incorrect, giving rise to cycles of social mistrust.

[12] Cole and Kocherlakota (2005) have questioned the role of memory in sustaining mutual cooperation in an infinitely repeated prisoner dilemma game with imperfect monitoring. They have examined the extent to which the set of equilibrium payoffs with infinite-memory strategies is a good approximation to the set of equilibrium payoffs with arbitrarily long finite-memory strategies. For some set of parameters, they have proved that defection in every period is the only strongly symmetric PPE, given bounded memory, regardless of the discount factor.

with certainty follow the equilibrium punishment rule. In the case of infinite memory, the punisher always has the appropriate incentives to follow the equilibrium strategies, independently of the enforced social norms. Such incentives stem from the fact that – by means of the information encoded in the history of signals – the next-period dynasty also voluntarily re-coordinates play on a cooperative path after one-period punishment. In the case of finite memory, we claim that social norms without SCI enforce generational autarky as a unique equilibrium, even when players have nearly total patience. Intuitively, by restricting the strategy profile (3) to one-period memory, which player defects cannot be determined. Indeed, the punisher has incentives to mimic cooperative behavior in order to generate a good signal with higher probability; similarly, the agent in the reward state is tempted to mimic the punisher in order to increase his current utility. Hence, defection in each period is the only strongly symmetric PPE.

Interestingly, when the social norm with SCI is operational, we can exploit the property of perfect observability of the y-action to show that a long run cooperative outcome can also be sustained under the restriction of one-period memory. To address this issue, we let agents exert three different levels of self-commitment actions: $y_t \in y \equiv \{0, \underline{y}, \overline{y}\}$ with $0 < \underline{y} < \overline{y}$. The following additional Assumption integrates the theoretical environment.

Assumption 4 (Payoff-Ranking Condition) *For each x_t and y_t, the following additional payoff-ranking condition holds:*

$$u\left(x, \underline{y}\right) > u\left(0, \overline{y}\right) \text{ and } u\left(0,0\right) + \delta\omega\left(0, \underline{y}\right) < u\left(0, \overline{y}\right) + \delta\omega\left(x, \overline{y}\right)$$

The first part of Assumption 4 implies that the cost of exerting $\overline{y}$ is sufficiently high to exceed the burden simultaneously generated by the cooperative action and the low-intensive self-commitment action. The second part implies that the benefits from cooperation, when old, are large enough to compensate for the cost inflicted by high-intensive self-commitment actions.

Let us consider the following one-period-memory variant of the strategy reported in (8):

$$\left(\theta^x\left(h^t\right), \theta^y\left(h^t\right)\right) = \begin{cases} \left(0, \overline{y}\right) & \text{if } \iota_t = 0 \\ \left(x, \underline{y}\right) & \text{otherwise.} \end{cases} \tag{13}$$

where $\left(\theta^x\left(\varnothing\right), \theta^y\left(\varnothing\right)\right) = \left(x, \underline{y}\right)$. For each $t \geq 1$, the flag-function ι_t is now as follows:

$$\iota_t = \begin{cases} 0 & \text{if } \left[z_t = B, y_{t-1} = \underline{y}, \mu_t \geq \mu\right] \text{ or } \left[y_{t-1} = 0, \forall z_t, \mu_t\right] \\ 1 & \text{if } \left[z_t = B, y_{t-1} = \underline{y}, \mu_t < \mu\right] \text{ or } \left[z_t = G, y_{t-1} = \underline{y}, \forall \mu_t\right] \\ 2 & \text{if } y_{t-1} = \overline{y}, \forall z_t, \mu_t \end{cases}$$

According to (13), strategies are contingent *only* on last-period observations, where the flag-function enables one to discern punishers from potential deviators with certainty. Specifically, $\iota_t = 2$ signals that in the previous period, a player has acted as punisher, whereas $\iota_t = 0$ warns of a potential deviator. Therefore, it suffices to check that: on the one hand, $\overline{y}$ is not so large as to discourage the punisher from following the prescribed punishment scheme; on the other hand, $\overline{y}$ is not so low as to

discourage the deviator from mimicking the punisher. The following Proposition yields the result.

Proposition 5 *Under one-period memory, if* $\mathcal{L} > \max\left\{1+\frac{u(0,\underline{y})-u(x,\underline{y})}{u(x,\underline{y})-u(0,\overline{y})}, \frac{u(0,\underline{y})-u(0,\overline{y})}{u(x,\underline{y})-u(0,\overline{y})}\right\}$*, then the best sustainable payoff enforced by social norms with SCI exceeds the reservation payoff.*
Proof *(See appendix).*

When external enforcement is sufficiently strong and players have one-period memory, Proposition 5 states that – by enlarging the self-commitment action space – social norms with SCI can successfully recover all the relevant information encoded in the full past history. This statement clarifies the role of history: *under finite memory and imperfect monitoring, only social norms with SCI enforce trust in equilibrium*. Hence, it reveals the *signal-driven* nature of social norms with SCI compared to the *history-driven* nature of social norms without SCI.

6.2 Self-Commitment Action as a Productive Activity

Heretofore, we have justified the existence of SCI as an unproductive device that credibly signals unobservable cooperative behavior and, in turn, fosters intergenerational trust. A natural question is whether our results hinge on the wastefulness of SCI. To investigate this question, we analyze a variant of the benchmark model in which generations exert *productive* y-actions. A prominent example is investment in children within a family. Every generation of parents not only makes myriad sacrifices for their offspring but also decides how much to invest in their well-being, for example, in terms of (public and private) education. Under perfect monitoring, Rangel (2003) has shown that at least a part of such investments is driven by strategic considerations: parents believe that investments in productive activity are the price they must pay to obtain the cooperative counterparts that they desire in old age (i.e., care, insurance, and status). Hence, the author concludes that x- and y-actions are strategic complements. Here, we show that under imperfect monitoring, this result remains *only* when y-actions play the role of self-commitment actions.

To address this issue, we require a small modification of the model. The key difference is that a *growth-enhancing* technology is now available. Using y_t as the sole input, the economy produces, with a one-period lag, a single homogenous good, f_{t+1}, according to the following technology:

$$f_{t+1} = \ell(y_t) \tag{14}$$

Assumption 5 (Productive SCI) $\ell : Y \to \mathbb{R}_+$ *is a strictly increasing function in* y_t *for each* t*, with* $\ell(0) \geq 0$.

The t-generation's intertemporal utility is now given by:

$$\upsilon(\boldsymbol{f}, \boldsymbol{x}, \boldsymbol{y}) = u(f_t, x_t, y_t) + \delta\omega(f_{t+1}, x_{t+1}, y_{t+1})$$

where $\boldsymbol{f} \equiv (f_t, f_{t+1})$ is the vector of outcomes of productive actions.

Assumption 6 (Preferences and Productive SCI) $u : \{\ell(0), \ell(y)\} \times X \times Y \to \mathbb{R}$ *and* $\omega : \{\ell(0), \ell(y)\} \times X \times Y \to \mathbb{R}$ *are increasing functions in* f_t*, for each* t*.*

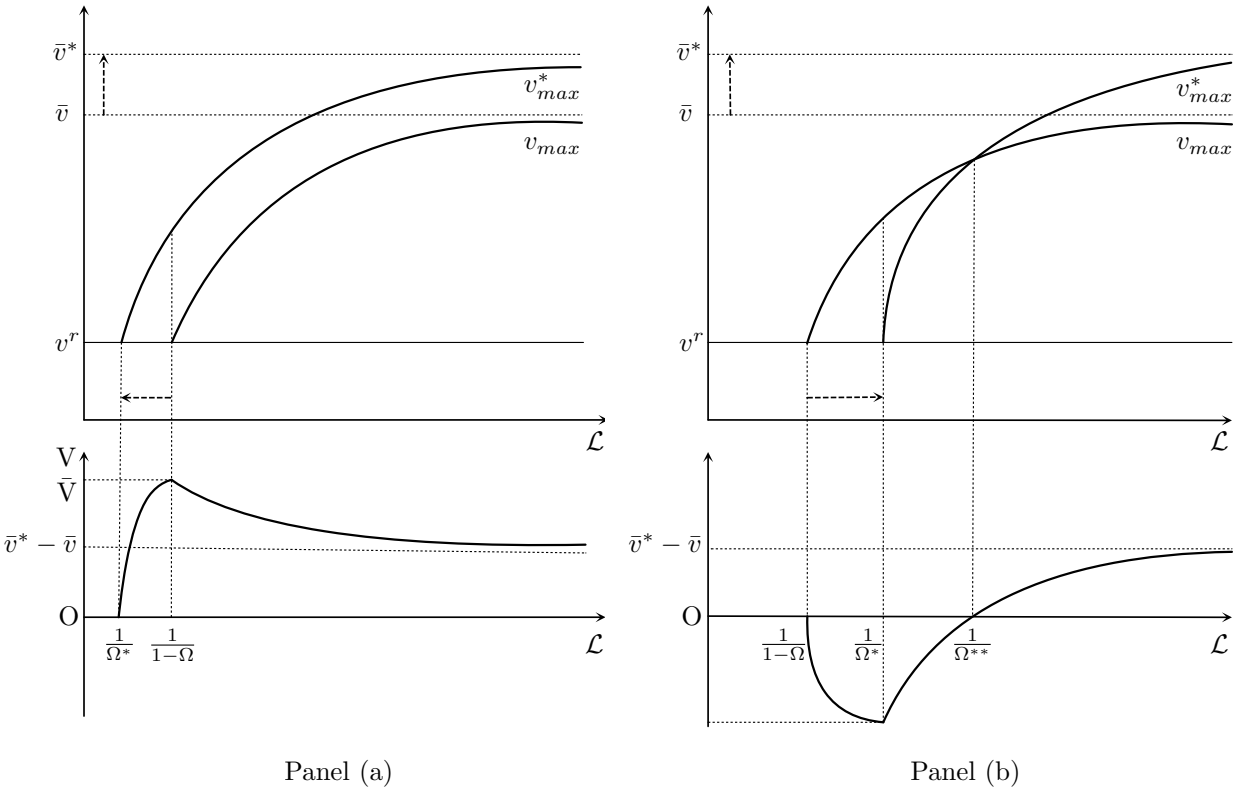
According to Assumption 6, current investment positively affects the payoffs of both next-period elderly agents (i.e., the investors) and the unborn young (i.e., the children of the investors). Hence, intertemporal utility, conditional on the provision of productive SCI, is strictly larger than intertemporal utility without investment. Similarly to the unproductive case, y_t is interpreted as a self-commitment action if the property of decreasing difference of current utility in (x_t, y_t) holds. Note that the expected cost of being cheated is necessarily larger when parents raise their children's productivity by investing in educational programs, i.e., $\omega(\ell(0), x, 0) - \omega(\ell(0), 0, 0) < \omega(\ell(y), x, y) - \omega(\ell(y), 0, y)$. Unlike unproductive SCI, the stationary efficient allocation implemented by a central planner with full commitment entails both $\boldsymbol{x}, \boldsymbol{y} \gg (0, 0)$. Replicating the analysis developed in section 4, the following Proposition holds.

Proposition 6 (Productive SCI) *Under productive SCI,* $V > 0$ *for any* $\mathcal{L} \in [1, \infty)$
Proof *(See appendix).*

This result is fundamentally different from the result articulated in Proposition 3. When SCI is productive, the effectiveness of the monitoring technology is not a critical element in the comparative analysis of different social norms: *social norms with SCI always Pareto dominate social norms without SCI*. Out of being growth-enhancing, productive SCI reduces gains from current deviations and increases losses from being cheated. These two implications unambiguously help signal strong intentions to cooperate along the unobservable dimension and, in turn, sustain the equilibrium strategic complementarity between the x- and y-actions.

Figure 6 in Panel (a) clearly illustrates the main properties of the equilibrium that follow from Proposition 6. Productive SCI has a twofold impact on the equilibrium outcome: (i) it *increases* the maximum sustainable payoff yield associated with better monitoring technology (technological effect), and (ii) it *expands* the parametric space of intergenerational cooperation (strategic effect). The former effect is measured by $\upsilon^*_{\max} - \upsilon_{\max}$ as $\mathcal{L}$ goes to infinity, whereas the latter effect is measured by the difference in agency costs, $\mathcal{C}(x, 0) - \mathcal{C}^*(x, y)$, which is positive for any degree of monitoring. Therefore, *productive SCI succeeds both in generating growth and in signaling cooperative behavior, for a given quality of external institutions.*

To highlight the robustness of our result, it is instructive to show how an alternative institution, called *Pseudo-Commitment Institution* (hereafter PCI) – although still prescribing the activation of observable, costly, and productive activities – can fail to reduce opportunistic behavior and thus fail to raise efficiency. If, contrary to Definition 2, $u(x_t, y_t)$ has nondecreasing difference in (x_t, y_t), then y_t would act as a *pseudo-commitment action*. In this case, current utility displays submodularity in (x_t, y_t) and accordingly reveals a preference for stronger substitutability between the two actions. In this scenario, productive activities can crowd out the cooperative counterpart, implying strategic substitutability between the two actions.



Panel (a) Panel (b)

Figure 6: Value of Productive SCI

Proposition 7 (Pseudo-Commitment Institution) *Under productive PCI, if* $V \leq 0$*, then* $\mathcal{I}(\mathcal{L}) \leq 0$*.*
Proof *(See appendix).*

Under productive PCI, agents – by exerting a costly y-action – only *pretend* to self-commit to mutually cooperative behavior, while in practice they signal a stronger intention to free-ride. Under sufficiently ineffective monitoring technology, the equilibrium strategic substitutability between the x- and y-actions weakens the benefits generated by the activation of the growth-enhancing technology – even if available. As a consequence, *the efficient policy prescription should encourage agents to comply with social norms that preclude growth*. Complementary to the analysis of productive SCI, Figure 6 in Panel (b) provides a graphical representation of the two opposite effects generated by PCI: (i) it *increases* the maximum sustainable payoff yield obtainable with a better monitoring technology, and (ii), in contrast to SCI, it *restricts* the parametric space of intergenerational cooperation. The magnitude of these two forces determines whether productive y-actions can be optimally activated along the equilibrium path. Specifically, if the institutions responsible for external enforcement are low-quality, i.e., $\mathcal{L} < \frac{1}{\Omega^{**}}$, optimal mechanisms require that individuals not coordinate on y-actions; whereas, if $\mathcal{L} > \frac{1}{\Omega^{**}}$, more efficient societies should activate growth-enhancing technologies. This result provides a straightforward explanation of the persistence of both low trust and underdevel-

opment in societies characterized by weak enforcement institutions. As North (1987) suggests, "*The inability of societies to develop effective, low cost enforcement is the most important source of both historical and contemporary underdevelopment in the Third World*".

Example II Consider a variant of the consumption-loan model described in Example I. Agents are born with $f_t \geq w$ units of wealth endowment. Thus, the individual budget constraints for young and old are modified to $c_t^1 = (f_t - y_t)(1 - \theta - x_t)$ and $c_{t+1}^2 = (f_{t+1} - y_{t+1})(\theta + x_{t+1})$, respectively. Without loss of generality, we adopt the functional form $f_{t+1} = w + \sqrt{y_t}$. Tables 4-6 decompose the SCI value into its technological, V^T, and strategic, V^S, components, and indicate the efficiency premium of SCI. Unlike in the previous example, there exists a threshold level $\hat{\gamma} \in \mathbb{R}_+$, such that if $\gamma > \hat{\gamma}$, then y_t is a self-commitment action. Note that Proposition 6 implies $\mathcal{I} > 0$ for any $\mathcal{L}$ and thus that social norms with SCI Pareto dominate social norms without SCI, for any feasible discount rate. A key insight of Table 4 is that an increase in intergenerational risk-aversion, $\frac{\eta}{\gamma}$, negatively affects SCI signaling power and, in turn, the value of SCI. Interestingly, this relationship is the opposite of the relationship stated in section 5 and provides a neat identification criterion. Clearly, the larger gain in efficiency associated with a lower degree of intergenerational risk-aversion is mainly absorbed by the strategic component of the SCI efficiency premium. The impact of policy interventions – such as wealth subsidies or mandatory transfers – on the signaling power of SCI is the reverse of the results achieved under unproductive SCI. To illustrate, compare Tables 2-3 to Tables 5-6.

Table 4: Value of Productive SCI, w=1, θ=0, η=10 and δ=0.98

γ	$\hat{\gamma}$	$\mathcal{I}$	$\overline{\mathcal{L}}$	$V(\overline{\mathcal{L}})$	$\frac{V^T(\overline{\mathcal{L}})}{v_{\max}}$	$\frac{V^S(\overline{\mathcal{L}})}{v_{\max}}$
1.4	1.203	0.008	1.298	0.100	0.063	0.019
1.7	1.203	0.017	1.276	0.135	0.067	0.049
2	1.203	0.024	1.256	0.167	0.069	0.080

Table 5: Value of Productive SCI, θ=0, γ=2, η=10 and δ=0.98

w	$\hat{\gamma}$	$\mathcal{I}$	$\overline{\mathcal{L}}$	$V(\overline{\mathcal{L}})$	$\frac{V^T(\overline{\mathcal{L}})}{v_{\max}}$	$\frac{V^S(\overline{\mathcal{L}})}{v_{\max}}$
0.8	1.470	0.021	1.288	0.163	0.088	0.049
1	1.203	0.024	1.256	0.167	0.069	0.080
1.2	1.018	0.025	1.221	0.171	0.055	0.104

Table 6: Value of Productive SCI, w=1, γ=2, η=10 and δ=0.98

θ	$\hat{\gamma}$	$\mathcal{I}$	$\overline{\mathcal{L}}$	$V(\overline{\mathcal{L}})$	$\frac{V^T(\overline{\mathcal{L}})}{v_{\max}}$	$\frac{V^S(\overline{\mathcal{L}})}{v_{\max}}$
0	1.203	0.024	1.256	0.167	0.069	0.080
0.01	1.219	0.022	1.298	0.153	0.075	0.074
0.02	1.237	0.019	1.350	0.140	0.081	0.068

Remarkably, an increase in wealth subsidies has an ambiguous effect on welfare. On the one hand, it reduces the value of SCI in its technological component, as it reduces the margin of productive improvement. On the other hand, it increases the value of SCI in its strategic component. In our specific parametric case, Table 5 documents a positive correlation between $V(\overline{\mathcal{L}})$ and w. Similar

arguments hold for policy interventions prescribing mandatory transfers (see Table 6).

Table 7: Productive PCI, w=1, θ=0, η=10 and δ=0.98

γ	$\hat{\gamma}$	$\mathcal{I}$	$\tilde{\mathcal{L}}$
0.5	1.203	−0.017	1.362
0.6	1.203	−0.016	1.307
0.7	1.203	−0.014	1.252

As final evidence, Table 7 provides a parametric configuration that fits the case of PCI. Indeed, for $\gamma < \hat{\gamma}$, the current payoff exhibits increasing difference in (x_t, y_t). In this circumstance, agents, by exerting y-actions, foster growth as well as opportunistic behavior. A negative value of $\mathcal{I}(\mathcal{L})$ warns of this possibility. Therefore, if $\mathcal{L} < \tilde{\mathcal{L}}$, efficiency requires that growth-enhancing technologies not be activated, even if they are available. □

6.2.1 Endogenous Institutional Changes

As highlighted in Corollary 2, when self-commitment actions are unproductive, the value of exchange crowds out the value of SCI. Conversely, the previous section shows that productive SCI enlarges the possibilities of exchange. Thus, a natural and challenging question is: Does growth-enhancing SCI generate a positive (negative) feedback loop that reinforces (undermines) its own viability? According to Greif (2002b), an institution is self-reinforcing when the process it entails – through its impact on the economic and political environment – increases the range of situations in which the institution is self-enforcing. On the other hand, an institution is self-undermining when it cultivates the seeds of its own decline by restricting the condition of self-enforceability. Interestingly, *depending on initial conditions*, productive SCI can be both self-reinforcing and self-undermining. On the one hand, if SCI were initially implemented because of both its strategic and technological effects, then the endogenous intensification of exchange due to growth would magnify the institutional self-enforceability condition. On the other hand, suppose that self-commitment actions were initially implemented only because of their strategic valence – this is the case when the investment cost exceeds the benefits of growth. In the long run, the institution gradually alters the environment in which agents interact: future generations indirectly benefit from the growth of trade, while at the same time, the growth of exchange undermines the initial signaling power of SCI. Thus, the endogenous growth process causes previously self-enforcing cooperative institutions to cease being so, possibly making alternative institutions desirable from a welfare perspective. Such a mechanism may illuminate certain trends in human history: (i) differences in human capital accumulation and, correspondingly, per-capita income across countries,[13] and (ii) the puzzling persistence of institutional outcomes, noted in Acemoglu, Johnson, and Robinson (2001).[14] The following example illustrates self-undermining SCI.

Example III Let agents be endowed with f_{t+1}: an endogenous function of both y_t exerted by the

[13]For an explanation based on technologically-driven incentives, see Galor and Zeira (1993).

[14]Countries and regions that were ruled centuries ago by despotic governments, or where powerful elites exploited uneducated peasants or slaves, today are plagued by institutional and organizational failure.

previous generation and a fraction $(1-\kappa)$ of the parental endowment. Without loss of generality, let it exhibit the functional form $f_{t+1} = \sqrt{y_t} + f_t\,(1-\kappa)$, with $f_0 = 1$ as the initial condition.

Table 8: Self-undermining Institution, f_0=1, θ=0, γ=0.6, η=10, δ=0.98 and κ=0.2

t	$\hat{\gamma}$	$\mathcal{I}$	$\overline{\mathcal{L}}$	$\tilde{\mathcal{L}}$	$V\left(\overline{\mathcal{L}}\right)$	$\frac{V^T\left(\overline{\mathcal{L}}\right)}{v_{\max}}$	$\frac{V^S\left(\overline{\mathcal{L}}\right)}{v_{\max}}$
0	1.400	0.010	1.213	·	0.023	−0.015	0.030
1	1.318	−0.001	∞	1.32	0.006	0.006	0

At $t = 0$, the first generation pays a cost $y_0 > 0$ and, when old, gains a return from productive investment by means of intergenerational transfers. As the investment cost exceeds the benefits of growth, the technological gain from SCI turns out to be negative, as reported in Table 8. Nevertheless, SCI is socially desirable because agents can exploit its signaling power when the likelihood ratio is sufficiently low. At $t = 1$, the second generation benefits from the implementation of SCI, both when young, as it earns a larger endowment, and when old, through more generous intergenerational transfers. In the specific parametric case, all parameters being equal, self-commitment actions foster future expected exchanges through technological improvement, but they also endogenously undermine the institution's viability. Indeed, when $\mathcal{L} < 1.32$, social norms with SCI make deviations more profitable, i.e., $\mathcal{I} < 0$, thereby failing to enforce cooperation. □

7 Applying the Model

The model is sufficiently general to apply to various contexts. In this section, we illustrate the mechanism proposed by providing various real-world examples. We refer to two prominent institutions within organizations: (i) *costly rituals* in religious organizations and (ii) *social responsibility* in corporate organizations. In each case, the evidence shows a clear link between trust, value of SCI, and contract enforceability. Specifically, they document how weak enforcement mechanism explain, in a fundamental way, the emergence and persistence of institutions acting as SCI.

Costly Rituals The role played by religion in increasing intra-group solidarity and cohesion has been widely documented by anthropologists. According to Durkheim (1912), individuals engage in collective rituals to bond with other community members. Rappaport (1979) regards ritual performance as a device for communication and resolution of collective action problems. Finally, Sosis (2002) views religious behavior as a *hard-to-fake* sign of commitment to facilitate intra-group cooperation. The economic literature has partially embraced the anthropological view of religious institutions. Beginning with the seminal paper by Iannaccone (1992), religion has been modeled as a club good characterized by positive returns to participatory crowding: club members benefit from access to an excludable good and an insurance network based on religiously motivated charitable acts. Iannaccone interprets time-intensive and wasteful activities – such as bizarre behavioral restrictions, sacrifice, stigma, and seemingly inefficient prohibitions – as devices to mitigate free riding problems and, in a heterogeneous society, screen for genuinely serious committed agents. By raising

the effective price of alternative activities, costly rituals induce both a substitution of resources from secular toward religious activities and an increase in individual welfare through indirect participatory crowding. In this paper, we defend a different argument. Like SCI, rituals are *voluntary, costly* – they entail both material costs and physical and psychological pain – and *perfectly observable*.[15] By reducing the marginal gain from opportunistic behavior with respect to imperfectly observable cooperative actions – such as mutual assistance or the local provision of public goods – they self-commit to members belonging to the same religious group. Furthermore, our theory suggests a fundamental link between external enforcement institutions adopted to monitor and publicize members' behavior and informal enforcement institutions as a key element of religious sects. Costly rituals are especially simple and powerful mechanisms for fostering social cohesion when monitoring difficulties impede punishment, exclusion, and other more standard solutions.

Practices of the Ultra-Orthodox Jewish community provide an illuminating illustration of the SCI theory at work. Israel's Ultra-Orthodox Jews are an intriguing and influential sect. Since 1948, the community has been pragmatically involved in state institutions. Its political support of governing coalitions has been rewarded with generous budgets. Interestingly, the education of most Ultra-Orthodox Jews focuses almost entirely on religious texts such as the *Torah* and *Talmud*, which they study full-time at *Yeshiva* – a religious school that provides no practical professional training. As a result, more than 60 percent of Ultra-Orthodox Jews live below the poverty line, compared with 12 percent of the non-Ultra-Orthodox Jewish population. These observations may suggest a causal relationship between Yeshiva attendance and poverty within the community, as years spent in Yeshiva could be spent accumulating valuable human capital. Such a relationship, however, appears hard to reconcile with the traditional utility-maximization theory. To rationalize Ultra-Orthodox practices, we require a theory of sacrifice. Intriguingly, a rational explanation of the emergence and persistence of such an institution is provided by the theory of SCI. In line with our modeling assumptions, religious observance requires (i) perfect observability - Yeshiva enrolment and full attendance - and (ii) costliness nature – Yeshiva attendance reduces productive time available to working age individuals. Furthermore, in a close community, collective action problems related to the use of common resources or the provision of local public goods might be especially severe owing to the absence of well-defined property rights and difficulties in monitoring individual contributions. In this context, *Yeshiva acts as an SCI: conformity with social norms that require Yeshiva attendance reduce agents' incentives to free-ride and, in turn, stimulate trust among group members.* In this perspective, wealth subsidies implemented to alleviate poverty turn out to be ineffective. As discussed in section 5, the communal response to the introduction of external subsidies is to further increase religious school attendance, which, in turn, dissipates the initial subsidy. Empirical evidence confirms our hypothesis. As documented in Berman (2000), population growth among Ultra-Orthodox Jews – which doubled from approximately 140.000 (3.6 percent of the Israel population) in 1979 to approximately 290.000 (5.1 percent of the Israel population) in 1995 – led to a remarkable increase in the community's po-

[15] Rappaport (1979) observes how ritual behaviors – although they appear to be shrouded in mystery – are deliberate and observable, with clear message to adherents. For example, the Jewish tradition recognizes the importance of observable actions in the famous dictum, "*We will do and we will understand*" (Exodus $24:7$). Judaism itself is fundamentally characterized as a religion of action – for example, performance of rituals or *mitzvoth* – rather than belief.

litical power. Such growth translated into stronger government support, particularly for Yeshiva students, notably since the first right-center coalition took power in 1977. In that period, among other reforms, Israel introduced military service exemption for full-time Yeshiva students attending the religious school until age 41 (or until age 35, for members of families with five children). In addition, in 1984, the Ultra-Orthodox party rapidly translated its political clout into dramatic increases in funding of its own system of schools. In line with our predictions, the conspicuous subsidies created a set of disincentives to work. Indeed, data clearly show that by the mid-1990s, men's labor force participation dropped to one-third, while the average duration of Yeshiva attendance lengthened. According to Berman (2000), in 1993-1996, approximately 46 percent of Ultra-Orthodox men aged 41-45 – that is, after the exemption had been granted – chose Yeshiva attendance over work.

Our theory also sheds new light on recent literature investigating, both theoretically and empirically, the economic implications of religious reforms.[16] We note especially the Reformation of the Catholic Church in Europe in the sixteenth-century. The forerunners of the Protestant Reformation were, both doctrinally and practically, united in their opposition to medieval Roman Catholic abuses. Among groups associated with the Radical Reformation, Calvinists were typically seen as the strictest in following rigorous discipline. Our theory provides a natural interpretation of the basic features of Calvinism, a doctrine that firmly rejected Catholic rituals and established effective new monitoring institutions to ensure proper Christian moral behavior. This development is in line with our main prediction of strategic substitutability between external enforcement and informal institutions acting as SCI (see Corollary 1). The hallmark of Calvinism was the rigorous enforcement of religious discipline.[17] To this end, Calvin, in his *Institutes of the Christian Religion*, proposed a stricter interpretation of the *Doctrines of Predestination*. God, who exercises absolute control over all creation, predestined the *Elect* for salvation and condemned the rest to damnation, without considerations of merits. Therefore, as neither good works nor church attendance can bring salvation, all rituals and adornments were prohibited.[18] To carry out the strict religious directives, a *Consistory* of church elders and ministers was established, with the aim of regulating the morals of members by monitoring and publicizing individuals' behavior. All aspects of people's lives, both public and private, were supervised to ensure proper Christian moral conduct, and anyone caught doing the devil's work was harshly punished. In the economic literature, a fascinating comparison of religious institutions is conducted by Levy and Razin (2012), who in addition (Levy and Razin, forthcoming) examine the determinants of Calvin's Reformation in Geneva. Using a simple two-period signaling model, they rationalize the Doctrines of Predestination and evaluate whether the introduction of the Consistory brought a welfare improvement. They argue that, when the benefits from spiritual utility are sufficiently large, the combination of theology and monitoring institutions increases community welfare.

[16] Among others, see Levy and Razin (2012), Guiso, Sapienza, and Zingales (2003), and Barro and McCleary (2006).

[17] For Calvinist Reformers, the integrity of the community depended upon exemplary conduct and behavior: "*Accordingly, as the saving doctrine of Christ is the soul of the church, so does discipline serve as its sinews, through which the members of the body hold together, each in its own place. Therefore, all who desire to remove discipline or to hinder its restoration…are surely contributing to the ultimate dissolution of the church*". [John Calvin, 1536, *Institutes of the Christian Religion*, $IV.xii.1$].

[18] In contrast to the Catholic Church, which implemented a heavy load of rituals in medieval times, Calvin rejected the seven sacraments, accepting only two sacraments as valid - Baptism and the Lord's Supper. In addition, Calvin encouraged his followers to return to the scriptures and read the Bible for themselves.

In contrast, our theory addresses a different aspect of the Calvinist Reform, namely, the apparent substitutability between costly religious rituals and the empowerment of the Consistory. As discussed above, we conjecture that the reduction of time-intensive ritual was the optimal communal response to the implementation of a powerful external institution. Increased trust is enforced without any need for costly signaling.

Social Responsibility In recent decades, issuance of environmental sustainability reports and engagement in *Corporate Social Responsibility* (hereafter CSR) activities have become increasingly prominent corporate practices.[19] Interestingly, while CSR was almost unknown in pre-globalization society, it is gaining momentum in globally integrated economies. Various factors have contributed to this trend: (i) increasing costs associated with global warming and public awareness of it, (ii) increased volatility of financial markets related to easier access to company information, (iii) the larger scope for cooperation in relation to environmental and social issues. The growing importance of those developments has led both governments and academics to explore how corporate performance might be affected by CSR practices and how they may represent an optimal communal response to market and public redistributive failures. Following McWilliams and Siegel (2001), CSR is defined as a situation where "*firms go beyond compliance and engage in actions that appear to further some social good, beyond the interests of the firms and that which is required by law*". Although numerous alternative definitions have been provided in the management literature, most scholars converge on the identification of two fundamental traits: (i) CSR activities aim to fulfill social goals at some cost to firms' profits, that is, the firm engages in costly and socially valuable activities that go beyond firm's legal obligations; (ii) corporations undertake CSR practices *voluntarily*. A large variety of activities meet these requirements.[20] Theoretically, rationalizing firms' private incentives to act in socially responsible ways remains an open issue. Evaluations of the costs and benefits of socially responsible behavior in previous literature suggest that CSR can arise from various motivations, including altruism and moral concern but also self-interest and threats by activists.[21] For example, a social entrepreneur – with the goal of altruistically redistribute corporate profits to social causes – may have an incentive to take over a profit-maximizing firm and convert it to a CSR firm at a financial loss. This interpretation presumes a desire by management to engage in philanthropy, an objective criticized by Friedman (1970): "*there is one and only one social responsibility of business–to use it(s) resources and engage in activities designed to increase its profits*".[22] In this latter view, being a good CSR corporate organization actually makes the firm more profitable. Indeed, consumers who perceive social giving positively may reward CSR firms in the marketplace by increasing demand for their products, as Baron (2001) argues. Alternatively, a firm could self-regulate by investing in CSR to avoid external pressure due to public regulation, as in Maxwell, Lyon, and Hackett (2000), or interest groups and activists, as in

[19] According to an ICCA global report survey (2010), approximately 30 percent of Fortune 500 companies have separate CSR departments.

[20] As illustrative examples: a credit institute might approve larger loans to poor borrowers than is required by the Authority governing lending practices in the financial sector; a manufacturer may achieve high-standard environmental performance through recycling and pollution abatement, significantly in excess of government requirements.

[21] See, among others, Baron (2001 and 2010) and Benabou and Tirole (2010).

[22] Friedman, M., 1970, The Social Responsibility of Business Is To Increase Profit, *New York Times Magazine*, pp. 32-33.

Baron (2010). Our theory suggests a view of social responsibility in organizations from a novel angle. CSR as SCI is a voluntary, costly and easily observed action. We argue that highly volatile financial markets and complex and interconnected production processes – outcomes of globalization – make it difficult for investors to evaluate management performance. This translates into lower rewards for firms in the marketplace, for example, through a reduction in shareholder entrenchment in badly behaving firms. According to Proposition 3, investment in CSR can be interpreted as an attempt to clear this financial hurdle. *If CSR practices were sufficiently costly to reduce gains from unobservable deviations, then they would foster goodwill and trust among investors and, in turn, protect the company in case of negative events.* A conspicuous empirical literature has struggled to quantitatively evaluate the link between CSR and some indicators of corporate financial performance.[23] Our prediction is consistent with most of the evidence, which finds that shareholders of firms with stronger environmental performance react less negatively to announcements of eco-harmful corporate behavior. In contrast, as shown by Flammer (2013), U.S. publicly traded companies that are subject to more severe environmental concerns may experience more dramatic stock price decreases upon announcement of eco-harmful events. Additionally, Godfrey, Merrill, and Hansen (2009) have shown that negative stock market reactions to announcements of legal actions against companies are significantly mitigated if firms undertake institutional CSR activities. Remarkably, our theory is also consistent with a negative relationship between CSR and corporate financial performance, that is, when CSR investments, such as PCI, are not sufficiently costly to deter opportunistic behavior and free riding. This finding might explain the apparent inconsistency of the results across different empirical analyses, as noted by McWilliams and Siegel (2001), and represents a prediction worth testing in future studies.

Let us conclude by examining historically more distant episodes. Greif (2002a and 2002b) has comprehensively shown how different types of social responsibility institutions spread widely throughout Europe as early as the eleventh and twelfth centuries. In that epoch, the state had no an effective authority and courts had weak enforcement power.[24] Despite this evidence, historical records document the proliferation of intercommunity impersonal transactions and a subsequent rapid development of long-distance trade. A suggestive explanation for this puzzle is that institutions of social responsibility were the driving force in the expansion of intercommunity impersonal exchanges. Greif (2002b) has focused on *Community Responsibility Systems* (hereafter CRS) in examining the expansion of trade in an epoch when cost (quid) and benefits (quo) were separated over time and space. He has observed how social heterogeneity and fractionalization prevented growth of trade in the early middle ages. In this scenario, only a socially homogeneous unit had could effectively punish members who cheated outsiders and, in turn, secure intercommunity exchange. In a CRS, traders exert a cost in acquiring a community affiliation. After becoming part of the community, members who declined to conform to internal conduct rules were excluded and their properties confiscated. Straightforwardly, CRS also resembles the mechanism embedded in SCI. Its costly and perfectly

[23] The *Socially Responsible Investing Studies* (www.sristudies.org) lists over 225 studies discussing elements of the CSR-corporate financial performance relationship.

[24] As documented by Hanawalt (1974), in well-organized societies such as England, there were few local courts with the power to enforce contracts. However, they were not immune to corruption, were mainly devoted to the defense of special local interests, and, in the final instance, were ineffective.

observable nature makes it an effective means for promoting cooperation when intercommunity exchange cannot rely on effective legal enforcement, as noted in Corollary 1. Historical evidence also suggests, paradoxically, that the system of communal responsibility contributed to its own demise. According to Greif (2002b), *"Ironically, it was the same process that the CRS fostered – processes through which trade expanded [...] – that reduced the economic efficiency and the intra-community political viability of CRS"*. During the late thirteenth century, various city-states in Italy contracted to abolish CRS, whereas in France and England, CSR was declared illegal and gradually replaced by a system based on individual responsibility. Which factors led to the disappearance of CRS? The claim regarding institutional change with productive SCI reported in section 6.2.1 provides a consistent explanation of the historical evidence. When SCI fosters growth and thus raises the expected future value exchanges, it also endogenously undermines its economic efficiency and political viability. Indeed, growth – by dampening the signaling power of self-commitment actions – makes SCI an ineffective commitment device. It even implies that, under weak monitoring institutions, enforcement of SCI is counterproductive in fostering trust and goodwill. Our model provides a theoretical explanation for the historical abolition of CRS. By expanding the scope and value of trade, CRS intensified rather than solved the moral hazard problems associated with long-distance trade.

8 Conclusions

In this study, we provide a theory of cooperation with Self-Commitment Institution. The power of our mechanism lies in the ability of short-lived agents to credibly signal their intention to cooperate by taking self-commitment actions. In this institutional environment, the sacrifices suffered by each player foster goodwill and trust among future generations and enlarge the scope of cooperation. Our theoretical framework applies to a variety of settings characterazed by two basic features: (i) that self-commitment actions are costly, fully observable, and satisfy a decreasing difference property over unobservables, and (ii) there is a prospect of follow-up intergenerational agreements, where such a prospect serves as a disciplining device, enforcing intergenerational trust.

Our main findings are the following: (i) when agents are sufficiently patient, players that conform to social norms with SCI are more willing to cooperate even after the realization of a negative event; thus, higher ex-ante efficiency is supported in equilibrium; (ii) intergenerational cooperation is more easily sustained in the presence of weak (strong) monitoring institutions and players who comply with social norms with (without) SCI; (iii) for any level of monitoring, the larger is the value of future exchanges, the lower is the value of SCI; (iv) bounded memory does not preclude the enforceability of trust under SCI; (v) productive SCI Pareto dominates social norms without SCI, for any quality of external enforcement.

The mechanism is unusual in several respects. First, it is simple. Second, it is seemingly counterproductive, as it relies on the voluntary sacrifice of short term benefits. Third, both theory and case studies suggest that the mechanism works in seemingly disparate social and economic circumstances (including religious sects, families, communes, social movement organizations, firms, and so on). Lastly, SCI does not require that individual cooperative contributions be accurately observed. It

is therefore a viable device to mitigate free riding when moral hazard is a relevant concern.

We should emphasize that the model takes the monitoring of organizations as exogenously given. The question of what determines the effectiveness of external enforcement institutions in an ongoing economy is an important one, but lies beyond the scope of this study. We view it as a further step toward a more comprehensive theory of cooperation. Lastly, the extension of our theory to an environment with heterogeneous players and random matching, as in early contributions to the institutional economics literature, may yield richer insights.

9 Appendix

Proof of Proposition 3: The decreasing difference property of $u(x_t, y_t)$ in (x_t, y_t) implies $V_{\mathcal{L}} < 0$ and $V_{\mathcal{L}\mathcal{L}} > 0$, for each $\mathcal{L}$. To prove this claim, suppose, for contradiction, that $\mathcal{I}(\mathcal{L}) \leq 0$ and thus that $\frac{u(0,y)-u(x,y)}{u(0,0)-u(x,0)} \geq \frac{\omega(x,y)-\omega(0,y)}{\omega(x,0)-\omega(0,0)}$. Simple algebra shows that necessarily $\frac{1}{1-\Omega} < \frac{1}{\Omega^*}$. This implies that the value of SCI is equal to $V(\mathcal{L}) = -\upsilon_{\max}$, for $\mathcal{L} \in \left[\frac{1}{1-\Omega}, \frac{1}{\Omega^*}\right)$. Furthermore, under Assumption 1, $\lim_{\mathcal{L}\to\infty} V(\mathcal{L}) = u(x,y) + \delta\omega(x,y) - (u(x,0) + \delta\omega(x,0))$ is strictly negative. Exploiting the monotonicity property of $V(\mathcal{L})$, we conclude that if $\mathcal{I}(\mathcal{L}) \leq 0$, then $V(\mathcal{L}) < 0$, for each $\mathcal{L}$.

Proof of Proposition 4: It is straightforward to show that Conditions 2, 6, and 7, are simultaneously satisfied in the interval $\delta \in [\underline{\delta}, 1)$, with

$$\underline{\delta} \equiv \frac{u(0,0) - u(0,y)}{(\omega(x,y) - \omega(0,y)) - \frac{u(0,y)-u(x,y)}{u(0,0)-u(x,0)}(\omega(x,0) - \omega(0,0))}$$

Hence, we conclude that $\Pi \equiv \{(\pi_0, \pi_x) \in [0,1] \times [0,1] \mid \text{Conditions 2, 6, and 7}\} \neq \varnothing$, when $\delta \geq \underline{\delta}$.

Proof of Corollary 1: Let $\overline{\mathcal{L}}$ denotes the likelihood ratio corresponding to the highest feasible value of SCI, i.e., $\overline{V} \equiv V(\overline{\mathcal{L}}) > V(\mathcal{L})$, for each $\mathcal{L}$. As, under decreasing difference, $V_{\mathcal{L}} < 0$, then it is necessarily the case that $\overline{\mathcal{L}} < \mathcal{L}$, for any feasible likelihood ratio that belongs to the space Π.

Proof of Proposition 5: The three-state automaton representation of strategy (13) is structurally similar to the strategy reported in (8). In the reward and punishment states, the intertemporal payoff functions are equal to, respectively:

$$\upsilon_{\phi_R} = u(x, \underline{y}) + \delta\{\omega(x, \underline{y}) - \pi_x(1-\mu)(\omega(x, \underline{y}) - \omega(0, \overline{y}))\}$$

and

$$\upsilon_{\phi_{P_\tau}} = u(0, \overline{y}) + \delta\omega(x, \underline{y})$$

for each $\tau = \{1, 2\}$. We first must check that in the reward state, players have no incentive to deviate. Therefore, we control for the following set of enforceability constraints:

$$\begin{aligned}
\upsilon_{\phi_R} &\geq u(0, \underline{y}) + \delta\{\omega(x, \underline{y}) - \pi_0(1-\mu)(\omega(x, \underline{y}) - \omega(0, \overline{y}))\} && (15)\\
&\geq u(0, \overline{y}) + \delta\omega(x, \underline{y}) && (16)\\
&\geq u(0,0) + \delta\omega(0, \overline{y}) && (17)\\
&\geq u(x, \overline{y}) + \delta\omega(x, \underline{y}) && (18)\\
&\geq u(x,0) + \delta\omega(0, \overline{y}) && (19)
\end{aligned}$$

Second, in the punishment state, the following multi-side enforceability conditions must hold:

$$
\begin{aligned}
v_{\phi_{P_\tau}} &\geq u\left(0,\underline{y}\right)+\delta\left\{\omega\left(x,\underline{y}\right)-\pi_0\left(1-\mu\right)\left(\omega\left(x,\underline{y}\right)-\omega\left(0,\overline{y}\right)\right)\right\} && (20)\\
&\geq u\left(x,\underline{y}\right)+\delta\left\{\omega\left(x,\underline{y}\right)-\pi_x\left(1-\mu\right)\left(\omega\left(x,\underline{y}\right)-\omega\left(0,\overline{y}\right)\right)\right\} && (21)\\
&\geq u\left(0,0\right)+\delta\omega\left(0,\overline{y}\right) && (22)\\
&\geq u\left(x,\overline{y}\right)+\delta\omega\left(x,\underline{y}\right) && (23)\\
&\geq u\left(x,0\right)+\delta\omega\left(0,\overline{y}\right) && (24)
\end{aligned}
$$

Clearly, Eqs. (18), (19), and (24) are dominated by Eqs. (16), (17), and (22), respectively, and then disregarded. Furthermore, under Assumption 1, the inequality (23) is trivially satisfied. The one-period memory strategy (13) is a PPE only if (i) from Eqs. (16) and (21), $\mu=1-\frac{u\left(x,\underline{y}\right)-u\left(0,\overline{y}\right)}{\delta\pi_x\left(\omega\left(x,\underline{y}\right)-\omega\left(0,\overline{y}\right)\right)}$, (ii) Eq. (16) dominates both Eqs. (15) and (17), and (iii) Eq. (21) dominates both Eqs. (20) and (22). Assumption 4 ensures that the deviation $u\left(0,\overline{y}\right)+\delta\omega\left(x,\underline{y}\right)$ is individually preferable to the deviation $u\left(0,0\right)+\delta\omega\left(0,\overline{y}\right)$. Algebraic manipulations of the remaining conditions yield $M\equiv\left[\underline{\mu},\overline{\mu}\right]$ as a feasible randomization range, with $\overline{\mu}\equiv 1-\max\left\{\frac{u\left(0,\underline{y}\right)-u\left(0,\overline{y}\right)}{\delta\pi_0\left(\omega\left(x,\underline{y}\right)-\omega\left(0,\overline{y}\right)\right)},\frac{u\left(0,\underline{y}\right)-u\left(x,\underline{y}\right)}{\delta(\pi_0-\pi_x)\left(\omega\left(x,\underline{y}\right)-\omega\left(0,\overline{y}\right)\right)}\right\}$and $\underline{\mu}\equiv 1-\frac{\left(u\left(x,\underline{y}\right)+\delta\omega(x,\overline{y})\right)-\left(u(0,0)+\delta\omega\left(0,\underline{y}\right)\right)}{\delta\pi_x\left(\omega\left(x,\underline{y}\right)-\omega\left(0,\overline{y}\right)\right)}$. As last step, to guarantee feasibility, i.e., $\mu\in M$ with $M\neq\varnothing$, it suffices for $\mathcal{L}$ to be sufficiently large, namely, $\mathcal{L}>\max\{1+\frac{u\left(0,\underline{y}\right)-u\left(x,\underline{y}\right)}{u\left(x,\underline{y}\right)-u\left(0,\overline{y}\right)},\frac{u\left(0,\underline{y}\right)-u\left(0,\overline{y}\right)}{u\left(x,\underline{y}\right)-u\left(0,\overline{y}\right)}\}$.

Proof of Proposition 6: Under Assumption 6, $\lim_{\mathcal{L}\to\infty}V\left(\mathcal{L}\right)=u\left(\ell\left(y\right),x,y\right)+\delta\omega(\ell(y),x,y)-(u(\ell(0),x,0)+\delta\omega\left(\ell(0),x,0\right))>0$. Furthermore, as SCI is productive, then the expected cost of being cheated is increasing in y_t, i.e., $\omega\left(\ell\left(0\right),x,0\right)-\omega\left(\ell\left(0\right),0,0\right)<\omega(\ell\left(y\right),x,y)-\omega(\ell\left(y\right),0,y)$. The decreasing difference property of $u\left(\ell\left(y_t\right),x_t,y_t\right)$ in (x_t,y_t) requires $u\left(\ell\left(y\right),0,y\right)-u\left(\ell\left(y\right),x,y\right)<u\left(\ell\left(0\right),0,0\right)-u\left(\ell\left(0\right),x,0\right)$. The latter two conditions jointly imply $\frac{u(\ell(0),0,0)-u(\ell(0),x,0)}{\omega(\ell(0),x,0)-\omega(\ell(0),0,0)}>\frac{u(\ell(y),0,y)-u(\ell(y),x,y)}{\omega(\ell(y),x,y)-\omega(\ell(y),0,y)}$, and thus $\mathcal{I}\left(\mathcal{L}\right)>0$, for each $\mathcal{L}$. Replicating the argument used in the proof of Proposition 3: because $\frac{1}{1-\Omega}>\frac{1}{\Omega^*}$, $V\left(\mathcal{L}\right)>0$, for each intensity of monitoring.

Proof of Proposition 7: The proof replicates the steps in the proof of Proposition 3. By increasing difference, $V_{\mathcal{L}}>0$ and $V_{\mathcal{L}\mathcal{L}}<0$, for each $\mathcal{L}$. To prove this claim, for contradiction, suppose $\mathcal{I}\left(\mathcal{L}\right)>0$ and, in turn, $\frac{u(\ell(0),0,0)-u(\ell(0),x,0)}{\omega(\ell(0),x,0)-\omega(\ell(0),0,0)}>\frac{u(\ell(y),0,y)-u(\ell(y),x,y)}{\omega(\ell(y),x,y)-\omega(\ell(y),0,y)}$. After some algebraic manipulations, we obtain the inequality $\frac{1}{1-\Omega}>\frac{1}{\Omega^*}$, which implies, jointly with Assumption 6, that $V\left(\mathcal{L}\right)>0$, for each $\mathcal{L}$.

References

[1] Abreu, D., Pearce, D., and E. Stacchetti, 1990, Toward a Theory of Discounted Repeated Games with Imperfect Monitoring, *Econometrica*, 58 (5), 1041-1063.

[2] Acemoglu, D., Johnson, S., and J.A., Robinson, 2001, The Colonial Origins of Comparative Development: An Empirical Investigation, *American Economic Review*, 91(5), 1369-1401.

[3] Acemoglu, D., and M., O, Jackson, 2011, History, Expectations, and Leadership in the Evolution of Social Norms, mimeo

[4] Ali, S., N., and D., A., Miller, 2013, Enforcing Cooperation in Networked Societies, mimeo.

[5] Anderlini, L., Gerardi, D., and R., Lagunoff, 2010, Social Memory, Evidence, and Conflict, *Review of Economic Dynamics*, 13, 559-574.

[6] Anderlini, L., and D., Terlizzese, 2012, Equilibrium Trust, mimeo.

[7] Banfield, E., 1958, The Moral Basis of a Backward Society, *Free Press*, New York.

[8] Baron, D., 2001, Private Politics, Corporate Social Responsibility and Integrated Strategy, *Journal of Economics and Management Strategy*, 10, 7-45.

[9] Baron, D. P., 2010, Morally-Motivated Self-Regulation, *American Economic Review*, 100(4), 1299-1329.

[10] Barro, R. J, and R. M., McCleary, 2006, Religion and Economy, *Journal of Economic Perspectives*, 20(2), 49-72.

[11] Benabou, R., and J., Tirole, 2010, Individual and Corporate Social Responsibility, *Economica*, 77, 1-19.

[12] Berman, E., 2000, Sect, Subsidy and Sacrifice: An Economists' View of Ultra-Orthodox Jews, *The Quarterly Journal of Economics*, 115 (3), 905-953.

[13] Bhaskar, V., 1998, Informational Constraints and the Overlapping Generations Model: Folk and Anti-Folk Theorems, *Review of Economic Studies*, 65(1), 135-149.

[14] Cole, L., and N., Kocherlakota, 2005, Finite memory and imperfect monitoring, *Games and Economic Behavior*, 53(1), 59-72.

[15] Cremer, J., 1986, Cooperation in Ongoing Organizations, *Quarterly Journal of Economics*, 100, 33-49.

[16] Dixit, A., K., 2003, Trade Expansion and Contract Enforcement, *Journal of Political Economy*, 111, 1293-1317.

[17] Durkheim, E., 1912, The Elementary Forms of Religious Life, *Free Press*, New York.

[18] Ellison, G., 1994, Cooperation in the Prisoner's Dilemma with Anonymous Random Matching, *Review of Economic Studies*, 61 (3), 567-88.

[19] Flammer, C., Corporate Social Responsibility and Shareholder Reaction: The Environmental Awareness of Investors, *Academy of Management Journal*, forthcoming.

[20] Fudenberg, D., Levine, D., and E., Maskin, 1994, The Folk Theorem in Repeated Games with Imperfect Public Information, *Econometrica*, 62, 997-1039.

[21] Galor, O., and J., Zeira, 1993, Income Distribution and Macroeconomics, *Review of Economic Studies*, 60(1), 35-52.

[22] Godfrey, P. C., Merrill, C. B., and J. M. Hansen, 2009, The Relationship Between Corporate Social Responsibility and Shareholder Value: An Empirical Test Of The Risk Management Hypothesis, *Strategic Management Journal*, 30, 425–445.

[23] Ghosh, P., and D., Ray, 1996, Cooperation in community interaction without information flows, *Review of Economic Studies*, 63 (3), 491-519.

[24] Greif, A., Milgrom, P., and B. R. Weingast, 1994, Coordination, Commitment, and Enforcement: The Case of the Merchant Guild, *Journal of Political Economy*, 102(4), 745-776.

[25] Greif, A., 2002*a*, Historical Institutional Analysis, *Cambridge University Press*.

[26] Greif, A., 2002*b*, Institutions and Impersonal Exchange: From Communal to Individual Responsibility, *Journal of Institutional and Theoretical Economics*, 158 (1), 168-204.

[27] Green, E., and R., Porter, 1984, Noncooperative Collusion under Imperfect Price Information, *Econometrica*, 52, 87-100.

[28] Guiso, L., Sapienza, P., and L., Zingales, 2003, People's Opium? Religion and Economic Attitudes, *Journal of Monetary Economics*, 50(1), 225-282.

[29] Hanawalt, B., 1974, The Peasant Family and Crime in Fourteenth-Century England, *Journal of British Studies*, 13, 1-18.

[30] Iannaccone, L. R, 1992, Sacrifice and Stigma: Reducing Free-Riding in Cults, Communes, and Other Collectives, *Journal of Political Economy*, 100(2), 271-291.

[31] Kandori, M., 1992*a*, Social Norms and Community Enforcement, *Review of Economic Studies*, 59(1), 63-80.

[32] Kandori, M., 1992*b*, Repeated Games Played by Overlapping Generations of Players, *Review of Economic Studies*, 59, 81-92.

[33] Kranton, R., E., 1996, Reciprocal Exchange: A self-Sustaining System, *The American Economic Review*, 86 (4), 830-851.

[34] Kreps, D. M., 1990, Corporate Culture and Economic Theory, in J. E. Alt and K. A. Shepsle, eds., *Perspectives on Positive Political Economy, Cambridge University Press*.

[35] Levy, G., and R. Razin, 2012, Religious Beliefs, Religious Participation and Cooperation, *American Economic Journal: Microeconomics*, 4(3), 121-151.

[36] Levy, G., and R. Razin, Calvin's Reformation in Geneva: Self and Social Signalling, *Journal of Public Economic Theory*, forthcoming.

[37] Maxwell, J. W., T. P. Lyon, and S. C. Hackett, 2000, Self-Regulation and Social Welfare: The Political Economy of Corporate Environmentalism, *Journal of Law and Economics*, 43(2), 583-617.

[38] Mailath, G., and L. Samuelson, 2006, Repeated Games and Reputations, *Oxford University Press*.

[39] McWilliams, A., and D., Siegel, 2001, Corporate Social Responsibility: A Theory of the Firm Perspective, *Academy of Management Review*, 26, 117-127.

[40] North, D., 1987, Institution, Transaction Costs and Economic Growth, *Economic Inquiry*, 25, 419-428.

[41] Putnam, R. D., 1993, Making Democracy Work, *Princeton University Press*.

[42] Rangel, A., 2003, Forward and Backward Intergenerational Goods: Why Is Social Security Good for the Environment?, *The American Economic Review*, 93(1), 813-834.

[43] Radner, R., 1986, Repeated Partnership Games with Imperfect Monitoring and No Discounting, *Review of Economic Studies*, 53 (1), 43-57.

[44] Rappaport, R., 1999, Ritual and Religion in the Making of Humanity, *Cambridge University Press*.

[45] Sosis, R., 2000, Religion and Intragroup Cooperation: Preliminary Results of a Comparative Analysis of Utopian Communities, *Cross-Cultural Research*, 34 (1), 70-87.

[46] Tabellini, G., 2008, The Scope of Cooperation: Values and Incentives, *The Quarterly Journal of Economics*, 123 (3), 905-950.

[47] Tabellini, G., 2010, Culture and Institutions: Economic Development in the Regions of Europe, *Journal of the European Economic Association*, 8(4), 677-716.

[48] Topkis, D., M., 1998, Supermodularity and Complementarity, *Princeton University Press*.