Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Save, strike, sustain: anti-cheating policy as leadership over an evolutionary public-goods game with hysteresis

Luca Meacci
Università degli Studi di Firenze
Francisco J. Muñoz
fcomunoz@alumnos.upm.es
Universidad Politécnica de Madrid
Juan Carlos Nuño
Universidad Politécnica de Madrid

Research Article

Keywords: evolutionary games, imitation dynamics, tax evasion, deterrence, Stackelberg leadership, optimal control, hysteresis

Posted Date: November 25th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11663526/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Save, strike, sustain: anti-cheating policy as leadership over an evolutionary public-goods game with hysteresis

Luca Meacci[1,2], Francisco J. Muñoz[3*], Juan Carlos Nuño[3]

[1]Dipartimento di Matematica e Informatica "Ulisse Dini", Università degli Studi di Firenze, Viale Morgagni, 67/a, 50134, Firenze, Italy.
[2]Dipartimento di Ingegneria dell'Informazione (DINFO), Università degli Studi di Firenze, Via di S. Marta, 3, 50139, Firenze, Italy.
[3*]Departamento de Matemática Aplicada, Universidad Politécnica de Madrid, 28040, Madrid, Spain.

*Corresponding author(s). E-mail(s): fcomunoz@alumnos.upm.es;
Contributing authors: luca.meacci@unifi.it; juancarlos.nuno@upm.es;

Abstract

A society can fight cheaters (individuals who benefit from public goods without contributing to them) by spending on deterrence or on education, and the classical question is how much to spend and how to split the budget. We revisit a population model of this problem as an evolutionary game. The original contagion dynamics is exactly an imitation (replicator-type) dynamics in which cheating has a constant payoff advantage, so policing removes cheaters but never changes their incentives. Adding a sanction turns deterrence into a payoff effect. The followers' dynamics is bistable: a compliant and a cheating state coexist for intermediate budgets, a feature present but unnoticed in the original model. Sanctions move the tipping point to smaller budgets, and allowing the rich to evade overturns the education-heavy recommendation unless sanctions are in place. We then let the government act as a leader that commits to a policy path while the population follows an imitation dynamics. Under a present-value budget, the numerically optimal policy from a collapsed state has three phases, save, strike and sustain, satisfies the conditions of Pontryagin's maximum principle, and raises discounted welfare 1.8 to 3.7 times over the best constant policy. In a spatial version the hysteresis band shrinks to a single Maxwell point near its lower edge, which a deterministic front theory predicts accurately.

Keywords: evolutionary games, imitation dynamics, tax evasion, deterrence, Stackelberg leadership, optimal control, hysteresis

MSC Classification: 91A22 , 91A65 , 49N90 , 91D10 , 37N40

JEL Classification: C73 , C61 , H26 , K42

1 Introduction

Public goods are financed by contributions that individuals are tempted to withhold. Tax evasion, benefit fraud and petty crime are instances of the same social dilemma: the cheater enjoys the common wealth without paying for it. Since Becker (1968) and Allingham and Sandmo (1972), the economic analysis of deterrence has asked how the probability and severity of punishment shape the decision to cheat. Game theory added the strategic side: in inspection games the inspector's effort and the inspectee's violation probability are jointly determined (Avenhaus et al. 2002), with the counter-intuitive result that, in mixed equilibrium, raising penalties changes the inspector's behaviour rather than the offender's (Tsebelis 1989; Andreozzi 2004). Evolutionary game theory studies the same dilemma in large populations, where punishment can stabilize cooperation in public-goods games (Brandt et al. 2006; Hauert et al. 2007), especially when it is organized as an institution funded by the contributors (Sigmund et al. 2010); spatial structure adds further effects (Szabó and Fáth 2007; Szolnoki et al. 2011; Perc et al. 2017).

A different line of work describes crime and cheating with population dynamics in the style of ecology. Nuño et al. (2008) proposed a predator–prey "triangle" of owners, criminals and guards; Nuño et al. (2011) called a society *criminal-prone* when its criminal-free steady state is unstable, and compared police size with social promotion through the stability of that state; and Nuño et al. (2010) (hereafter NHP) asked how much a society should invest against cheaters and how to split the investment between education and police. In the NHP model a poor class, a rich class and a class of cheaters evolve together with the total wealth w; the policy is a pair (ϕ, s), where ϕ is the fraction of wealth spent against cheating and s the share devoted to education. The stationary analysis showed that wealth is maximized at $(\phi, s) \approx (0.05, 0.9)$, an education-heavy policy. Related work on crime as a reaction–diffusion or agent-based phenomenon includes Short et al. (2008) and Berestycki and Nadal (2010), and the strategic defence of public goods against a thief was analysed as a two-target game by Bruni et al. (2013). The NHP model was later extended to a cellular automaton with local interactions (Meacci 2015).

Tax evasion in particular has a large agent-based literature, reviewed by Pickhardt and Prinz (2014). Taxpayers are placed on lattices or networks and copy the behaviour of their neighbours, either through Ising-type spin dynamics (Zaklan et al. 2009; Pickhardt and Seibold 2014) or through heterogeneous behavioural types with Allingham–Sandmo decisions (Bloomquist 2006; Korobow et al. 2007; Hokamp and Pickhardt 2010); the network topology changes aggregate compliance (Andrei et al.

2014). Antoci et al. (2014) study the replicator dynamics of cheaters, honest taxpayers and peer punishers. These models take audit probability and penalty as given. Here enforcement is financed from the wealth that compliance creates, which closes a fiscal feedback loop, and the government chooses enforcement as a leader.

The NHP model has no explicit incentives: cheaters are recruited by contagion at fixed rates, and police only removes them. This paper makes the strategic structure explicit and asks what changes when the government is treated as a player who commits to a policy over time. Our contributions are:

1. We show that the NHP dynamics is exactly a mutation–imitation dynamics of a public-goods game in which the payoff advantage of cheating is constant (Proposition 1). Adding a sanction F makes deterrence act on payoffs, and the NHP model is the special case $F = 0$.
2. We characterize the stationary states of the followers' dynamics. The dynamics is bistable for intermediate budgets, already for the parameters of NHP, contrary to the uniqueness stated there. We give a closed-form invasion threshold for the compliant state (Proposition 2): a sanction divides the detection rate needed for compliance by $1+\lambda F z_0$. Allowing the rich to evade shows that the education-heavy optimum is robust only if sanctions are in place.
3. We formalize the government as a Stackelberg leader over an evolutionary dynamics, with optimistic and robust values when the followers' response is multivalued. With a budget cap, a robust leader (one who expects the worst equilibrium) needs twice the budget of an optimistic one when police is congested. Over time, under a present-value budget, the numerically optimal open-loop policy from a collapsed society has a three-phase structure, *save, strike, sustain*, found from every starting point and confirmed by the maximum principle; it exploits hysteresis and raises welfare 1.8–3.7 times over constant policies.
4. In a spatial cellular-automaton version, fronts between compliant and cheating regions select a single threshold (a Maxwell point), predicted by a deterministic front theory, inside the bistable band, close to its lower edge, so spatial societies are more compliant than the mean-field model suggests.

Section 2 presents the model, Section 3 the equilibria of the followers' dynamics, Section 4 the leader's problem, Section 5 the spatial version, and Section 6 discusses the results. Numerical methods, sensitivity analysis and finite-population checks are in the appendices.

2 The model

2.1 The NHP population model

Let z, u and y denote the fractions of poor taxpayers, rich taxpayers and cheaters ($z + u + y = 1$), and w the total wealth normalized by the contribution of an all-poor population. Cheaters are recruited from the poor only and return to the poor class when they stop cheating. The NHP model (their eq. 3.15, with the rates in their

Table 1 Parameters (yearly time unit) and policy variables. Values are those of NHP unless stated.

Symbol	Meaning	Value
$\alpha_0,\ \alpha_1$	baseline promotion rate; effect of education spending	0.01, 1
β	relegation rate (rich to poor)	0.04
$\tau_0,\ \tau_1$	spontaneous recruitment; contagion (imitation) strength	0.01, 0.1
$\eta_0,\ \eta_1$	effect of education on recruitment	0, 0
μ	spontaneous drop-out of cheaters	0.01
k	police efficiency	1
$a,\ \theta$	contribution of the rich; ordinary expenses	16, 1
F	sanction on detected cheaters, in years of tax (this paper)	0–100
λ	imitation response to sanctions (this paper)	τ_1
$\phi,\ s$	spending share against cheating; education share	policy

eqs. 3.10–3.14) reads

$$\begin{aligned} \dot{z} &= -\alpha(w)\, z + \beta\, u - \dot{y}, \\ \dot{y} &= \big(\tau_0(w) + \tau_1(w)\, y\big)\, z - \mu\, y - k(1-s)\phi\, w\, y, \\ \dot{w} &= z + a\, u - (\theta + \phi)\, w, \end{aligned} \tag{1}$$

with upward mobility $\alpha(w) = \alpha_0 + \alpha_1 s\phi w$ (education promotes the poor) and recruitment rates $\tau_i(w) = \tau_i/(1+\eta_i s\phi w)$. Taxes are 1 per poor and a per rich individual; θw are ordinary expenses and ϕw the extra spending against cheating, of which a share s goes to education and $1-s$ to police. The last term in $\dot{y}$ removes cheaters at a per-capita detection rate $k(1-s)\phi w$. Table 1 lists the parameters, which are those of NHP.

2.2 The followers' game

Each poor individual chooses between paying taxes (T) and cheating (C). In a large population the benefit from the public good, financed by w, is received by both strategies and cancels in the payoff difference; an individual's own contribution has a negligible effect on w. What distinguishes the strategies is the tax saved, normalized to $g = 1$, against the expected sanction. If a detected cheater pays a sanction F (in years of tax) and detection occurs at the NHP rate $d(w) = k(1-s)\phi w$, the payoff advantage of cheating is

$$\Delta\pi(w) = \pi_C - \pi_T = \frac{g}{1+\eta_1 s\phi w} - F\, d(w), \tag{2}$$

where the first term allows education to lower the perceived gain, the role of η_1 in NHP.

Individuals revise their strategy by pairwise proportional imitation (Schlag 1998): a reviser meets a random individual of the other strategy and switches with probability

proportional to the payoff difference. In the mean-field limit this gives a replicator-type flow $\lambda\, yz\, \Delta\pi$ between poor taxpayers and cheaters (Hofbauer and Sigmund 1998; Sandholm 2010). Adding spontaneous exploration τ_0, drop-out μ and removal by the police, the cheater equation becomes

$$\dot{y} = \tau_0 z + \big(\tau_1 g(w) - \lambda F d(w)\big)\, y\, z - \mu y - d(w)\, y, \qquad g(w) = \frac{1}{1+\eta_1 s\phi w}, \tag{3}$$

while the equations for z and w are as in (1). With $\lambda = \tau_1$ the imitation term is exactly $\tau_1 \Delta\pi\, yz$, so τ_1 is the imitation rate per unit of payoff; λ is kept separate in the sensitivity analysis.

Proposition 1 *For $F = 0$ (and any λ), system (3) with the z and w equations of (1) coincides with the NHP model. Hence the NHP contagion term $\tau_1 yz$ is a replicator-type imitation flow driven by a constant payoff advantage of cheating, $\Delta\pi = g(w)$, which does not depend on deterrence spending.*

Proof With $F = 0$, (3) reduces to $\dot{y} = \tau_0 z + \tau_1 g(w) yz - \mu y - k(1-s)\phi w y$, and $\tau_1 g(w) = \tau_1(w)$ in the notation of NHP; with $\eta_0 = 0$, $\tau_0(w) = \tau_0$. The other equations are unchanged. □

The proposition separates two channels of deterrence. In NHP, police spending acts only *mechanically*, by removing cheaters at rate d. In the game it also acts on *incentives*: the imitation flow reverses, and cheaters start imitating taxpayers, when $Fd(w) > g(w)$. The government chooses (ϕ, s, F), and later the paths $(\phi(t), s(t), F(t))$; the population responds through (3). We refer to the population as the followers and to the government as the leader.

Modelling assumptions.

Four choices deserve comment. (i) *Linear imitation.* Proportional imitation is the payoff-improving rule with the least memory (Schlag 1998) and keeps the model polynomial, which is what makes Lemma 1 below possible; a logit (Fermi) rule changes the rates far from $\Delta\pi = 0$ but not the location of equilibria where $\Delta\pi$ changes sign. (ii) *Scale of F.* Payoffs are measured in units of the yearly tax of a poor individual, so F is a number of years of tax. What deters is the expected sanction per year, Fd, compared with the gain $g = 1$. In the compliant states of interest the detection rate is $d \approx 0.04$–0.07 per year (e.g. $d = 0.053$ at $\phi = 0.012$, $s = 0.5$), so cheating stops paying when $F \gtrsim 1/d \approx 14$–25. We take $F = 30$ as a sanction just above this deterrence threshold: an evader is caught on average after $1/d \approx 20$ years and loses about thirty years of tax, i.e. the arrears plus a surcharge of the order of 50%. This is high but of the order of recovered arrears plus penalties comparable to the evaded amount, when the whole evasion period can be recovered; statutes of limitation make actual monetary sanctions smaller, and F should then be read as including non-monetary costs (criminal liability, reputation). $F = 100$ is an upper bound; how the dynamic results depend on this cap is discussed in Section 4. $F = 10$, below the threshold, is shown in Figs. 1–2. Note that in the cheating state d is much smaller ($1/d \approx 150$ years), so

no sanction in this range makes cheating unprofitable there: sanctions act on compliance once it exists, which is the origin of their effect on the lower fold. (iii) *Budget as a share of wealth.* Following NHP, spending is ϕw; this couples enforcement to the tax base and is the source of the fiscal feedback. Section 4 also checks a budget fixed in money terms. (iv) *Time scale.* Rates are per year, as in NHP; with $\tau_1 = 0.1$ and $\mu = 0.01$ the relaxation times are decades, which is why the timing of policy matters. The identification $\lambda = \tau_1$ uses a single imitation rate for gains and for sanctions; Appendix B varies λ separately.

3 Equilibria of the followers' dynamics

3.1 Reduction and invasion threshold

Lemma 1 *For fixed (ϕ, s, F), the equilibria of (1)–(3) are in one-to-one correspondence with the zeros in $(0,1)$ of the scalar function $G(y) = \dot{y}\big(z(y), y, w(y)\big)$, where $z(y) = \beta(1-y)/(\alpha(w(y))+\beta)$ and $w(y)$ is the unique positive root of $(\theta+\phi)\,\alpha_1 s\phi\, w^2 + \big[(\theta+\phi)(\alpha_0+\beta) - (1-y)\,a\,\alpha_1 s\phi\big] w - (1-y)\big(a\alpha_0+\beta\big) = 0$.*

Proof $\dot{z} + \dot{y} = 0$ gives $\alpha z = \beta u = \beta(1-y-z)$, hence $z(y,w)$. Substituting into $\dot{w} = 0$ and clearing denominators gives the quadratic, whose constant term is negative, so it has exactly one positive root. □

We write $G(y; \phi, s, F)$ when the dependence on the policy matters. If $s\phi = 0$ the quadratic degenerates to a linear equation and $w(y) = (1-y)(a\alpha_0+\beta)/\big((\theta+\phi)(\alpha_0+\beta)\big)$. For $F = 0$ the zeros of G are those of the fourth-degree polynomial in NHP (their eq. C1); the lemma holds for any F and gives a robust numerical procedure (Appendix A).

In the limit $\tau_0 \to 0$ the cheater-free state $y = 0$ is an equilibrium, with z_0 and w_0 given by Lemma 1 at $y = 0$. Linearizing (3) at $y = 0$ gives:

Proposition 2 (Invasion threshold) *Let $\tau_0 = 0$ and $\eta_1 = 0$. The cheater-free state is stable against invasion by cheaters if and only if the detection rate at that state satisfies*

$$d_0 = k(1-s)\phi w_0 \;>\; d^* = \frac{\tau_1 z_0 - \mu}{1 + \lambda F z_0}. \tag{4}$$

Proof At $y = 0$, $\partial\dot{y}/\partial y = \tau_1 z_0 - \lambda F d_0 z_0 - \mu - d_0$, which is negative iff $d_0(1+\lambda F z_0) > \tau_1 z_0 - \mu$. Since $\dot{y}$ is proportional to y, the Jacobian at $y = 0$ is block triangular, and the (z, w) block has negative trace and determinant $(\alpha+\beta)(\theta+\phi) - \alpha_1 s\phi\, z_0(a-1)$, which is positive for all parameters considered here. □

Proposition 2 is the game-theoretic version of the criminal-proneness criterion of Nuño et al. (2011): a society is cheater-prone exactly when (4) fails. Two differences are worth noting. In Nuño et al. (2011) recruitment depends on wealth and the police enters the critical recruitment rate quadratically; here the threshold is linear in the detection rate, and the sanction enters as a multiplier. Equation (4) makes the

Table 2 Analytical invasion threshold ϕ^* (Proposition 2) and numerical bistable band $[\phi_{\mathrm{lo}}, \phi_{\mathrm{hi}}]$, for two values of τ_0.

F	s	ϕ^* ($\tau_0 \to 0$)	band, $\tau_0 = 0.001$	band, $\tau_0 = 0.01$ (NHP)
0	0.5	0.0081	$[0.0090, 0.0114]$	$[0.0111, 0.0129]$
0	0.9	0.0109	$[0.0120, 0.0152]$	$[0.0149, 0.0170]$
30	0.5	0.0050	$[0.0057, 0.0086]$	$[0.0076, 0.0099]$
100	0.5	0.0027	$[0.0032, 0.0054]$	$[0.0045, 0.0064]$

role of sanctions explicit: they divide the detection rate required for compliance by $1 + \lambda F z_0$. As in Nuño et al. (2011), where criminality could be avoided with either very low or very high social promotion, the effect of education is non-monotone: the threshold ϕ^* is smallest at intermediate shares (0.0096, 0.0081, 0.0086, 0.0109 for $s = 0.25, 0.5, 0.75, 0.9$ with $F = 0$), because education both promotes the poor out of the recruitment pool and diverts money from detection. For $\tau_0 > 0$ the transcritical point becomes the lower fold of the bistable band; Table 2 shows that the threshold ϕ^* defined by $d_0 = d^*$ approaches the numerical lower fold ϕ_{lo} as τ_0 decreases.

3.2 Bistability and the effect of sanctions

All numerical results below were obtained by locating the zeros of G, classifying their stability with the 3×3 Jacobian and refining fold curves by bisection; bistable states were confirmed by integrating from different initial conditions. The numerical model reproduces NHP (Table A1 in Appendix A).

Bistability in the NHP model.

With the NHP parameters ($F = 0$, $k = 1$) there is a region of the policy plane, $\phi \approx 0.011$–0.023 and $s \approx 0.17$–0.98 (exact limits below), where a compliant state (few cheaters, high wealth) and a cheating state coexist, separated by a saddle. At $(\phi, s) = (0.012, 0.68)$, for example, $y^* = 0.13$ ($w^* = 9.8$) and $y^* = 0.82$ ($w^* = 1.0$) are both stable. NHP report a unique steady state for this setup; the band is narrow and lies just above $\phi = 0.01$, the value used in their sensitivity analysis, which is probably why it was not detected. The mechanism is a fiscal feedback: more cheaters mean less revenue, less police and more cheating, reinforced by imitation.

Folds and cusps.

Lemma 1 reduces the bifurcation analysis to a scalar equation.

Proposition 3 *For fixed F, the boundary of the bistable region in the (ϕ, s) plane consists of the saddle-node points $G(y) = \partial_y G(y) = 0$, and the bistable region ends at cusp points where in addition $\partial_y^2 G(y) = 0$. At a saddle-node the Jacobian of (1)–(3) has a zero eigenvalue.*

Proof Along the curve $\{\dot{z} + \dot{y} = 0,\ \dot{w} = 0\}$ parametrized by y (Lemma 1), equilibria are zeros of G, and two of them coalesce exactly when $G = G' = 0$. Since $\det J$ is proportional

Table 3 Cusp points (y^*, ϕ, s) from $G = G' = G'' = 0$, and folds at $s = 0.5$ from $G = G' = 0$. "None": the band reaches $s = 0$.

F	lower cusp	upper cusp	ϕ_{lo} $(s = 0.5)$	ϕ_{hi} $(s = 0.5)$
0	(0.442, 0.01459, 0.174)	(0.538, 0.02252, 0.981)	0.01112	0.01288
30	none	(0.538, 0.02255, 0.986)	0.00760	0.00988
100	none	(0.539, 0.02259, 0.991)	0.00446	0.00640

to $G'(y)$ times the (nonzero) determinant of the (z, w) constraint block, $G' = 0$ gives a zero eigenvalue. The cusp is the codimension-two point where three equilibria coalesce, i.e. where G'' also vanishes; it is a non-degenerate cusp if, in addition, $\partial_y^3 G \neq 0$ and the transversality condition $\det\begin{pmatrix} \partial_\phi G & \partial_s G \\ \partial_y\partial_\phi G & \partial_y\partial_s G \end{pmatrix} \neq 0$ holds. □

Solving these conditions (Table 3) gives the exact ends of the band: for $F = 0$ the bistable region spans $s \in [0.174, 0.981]$, with cusps at $\phi = 0.0146$ and $\phi = 0.0225$; for $F \geq 30$ the lower cusp disappears ($s = 0$ is bistable) and the upper cusp hardly moves. The fold values agree with the numerical continuation to five digits. At every cusp the non-degeneracy conditions hold: $\partial_y^3 G = -0.398$ at the lower cusp ($F = 0$) and between -0.095 and -0.096 at the upper cusps, and the transversality determinant is -0.122, 0.035, 0.047 and 0.076 ($F = 0$ lower and upper, $F = 30$, $F = 100$), so the cusps are generic and persist under small changes of parameters.

Sanctions shift the tipping point.

The bistable region is bounded by two fold curves meeting at cusps (Fig. 1). As F increases the region moves to smaller budgets and its lower cusp descends to $s = 0$, so for $F \geq 30$ even purely police-based policies are bistable. The shift is largest for small s, because sanctions act only through the police share. For fixed s the cheater fraction is S-shaped in ϕ (Fig. 2): a society that collapses below ϕ_{lo} does not recover when the budget returns to its previous level, but only above ϕ_{hi}.

The stationary optimum.

Maximizing stationary wealth over (ϕ, s) gives $(0.059, 0.90)$ with $w^* = 14.33$ for $F = 0$, confirming NHP, and $(0.057, 0.92)$ with $w^* = 14.36$ for $F = 100$. The optimum lies far from the bistable region, so in this setting sanctions change where the danger zone is but not where the optimum is. The next extension shows that this conclusion is fragile.

3.3 Evasion by the rich

In NHP only the poor can cheat. We add a class v of rich cheaters. Taxes and sanctions are proportional to income, so the relative advantage of cheating is the same in both classes and drives imitation within each class; the rich differ in spontaneous evasion ($\tau_0^r = 0.001$), detectability ($d_r = \rho d$) and damage (a rich evader removes $a = 16$ units of revenue). Attractors were obtained by integrating the five-dimensional system from a compliant and from a cheating initial condition on a 61×41 grid of (ϕ, s).

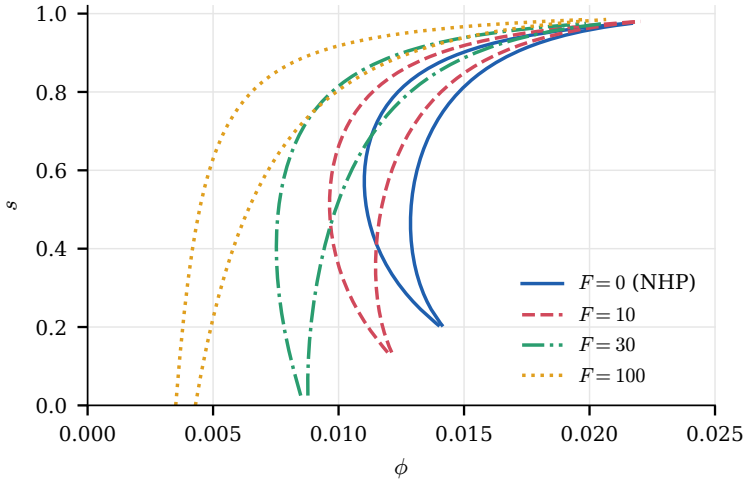

Fig. 1 Fold curves in the policy plane (ϕ, s) for several sanctions F. Each loop encloses the region with two stable equilibria.

An education-heavy policy creates a large rich class, and without sanctions evasion spreads by imitation within it (Fig. 3). At the NHP optimum $(0.05, 0.9)$ wealth falls from 14.3 to 6.1; the robust optimum moves to $(0.065, 0.775)$, towards police. With $F \geq 30$ the NHP optimum is restored ($w^* = 14.25$), and if rich evaders are harder to detect ($\rho = 0.5$) the optimal split moves towards police again.

Other extensions.

Two further mechanisms change the size of the bistable band. If police is congested, with detection $d = k(1 - s)\phi w/(1 + y/\kappa)$, the band widens strongly: at $s = 0.5$ it is $[0.0147, 0.0254]$ for $\kappa = 0.2$ and $[0.0168, 0.0305]$ for $\kappa = 0.1$, and it extends to $s = 0$. If education also lowers the gain from cheating ($\eta_1 = 100$), the band narrows to $[0.0072, 0.0078]$ and almost purely educational policies lose their hysteresis.

4 The government as a leader

This is a leader–follower problem of a particular kind. In a conventional Stackelberg game the followers solve an optimization problem and the leader anticipates their best response. Here the followers do not optimize: they imitate, and their response to a policy is the outcome of the evolutionary dynamics (1)–(3). The leader therefore optimizes over the trajectories, or the attractors, of that dynamics. This is the setting of Stackelberg evolutionary games, in which a rational leader steers an evolving population (Staňková et al. 2019; Stein et al. 2023; Kleshnina et al. 2023), and it is related to the optimal design of incentives in population games (Han and Tran-Thanh 2018; Stuckey et al. 2022). We make the ingredients explicit.

Leadership over an evolutionary dynamics.

Write $x = (z, y, w)$ and $\dot{x} = f(x, \pi)$ for (1)–(3) under policy π.

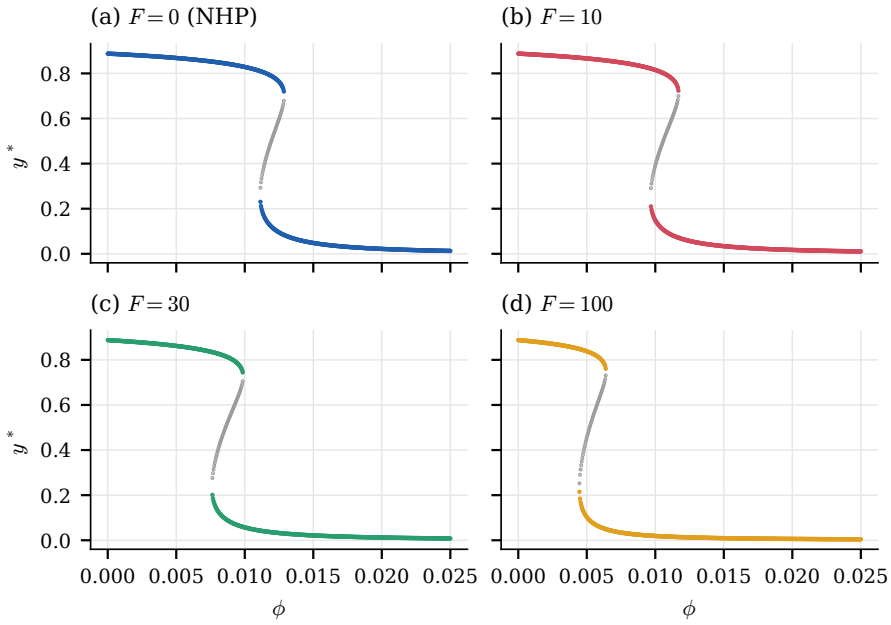

Fig. 2 Equilibrium cheater fraction versus ϕ at $s = 0.5$, $k = 1$. Colour: stable; grey: unstable (the threshold between basins).

1. *Commitment.* The leader chooses a policy before the followers move and cannot revise it: a constant $\pi = (\phi, s, F)$ in the static problem, an open-loop path $\pi(\cdot)$ in the dynamic one. The policy is common knowledge in the sense that it enters the payoffs (2).
2. *Information.* The leader knows the model, its parameters and, in the dynamic problem, the initial state x_0. Followers know only current payoffs and the strategies of the individuals they meet; they do not anticipate future policies.
3. *Response.* In the static problem the response to π is the set $\mathcal{A}(\pi)$ of asymptotically stable equilibria of $\dot{x} = f(x, \pi)$; which element is selected depends on the state the society starts from. In the dynamic problem the response is the unique solution $x(\cdot\,; \pi, x_0)$, since f is smooth and the state space is forward invariant.
4. *Optimality.* Let $W(x)$ be the leader's payoff. With several attractors two values are natural:

$$V^+(\pi) = \max_{x \in \mathcal{A}(\pi)} W(x), \qquad V^-(\pi) = \min_{x \in \mathcal{A}(\pi)} W(x). \tag{5}$$

An *optimistic* leader maximizes V^+ and a *robust* leader maximizes V^-. They are analogous to the strong (optimistic) and weak (pessimistic) Stackelberg solutions of bilevel optimization, in which the follower breaks ties among its best responses in favour of or against the leader (Loridan and Morgan 1996); here the multiplicity comes from coexisting attractors of the followers' dynamics rather than from ties among best responses. The difference $\max V^+ - \max V^-$ is the price of robustness. In the dynamic problem the response is unique and the leader maximizes $J[\pi] = J\big[x(\cdot\,; \pi, x_0), \pi\big]$, an optimal control problem.

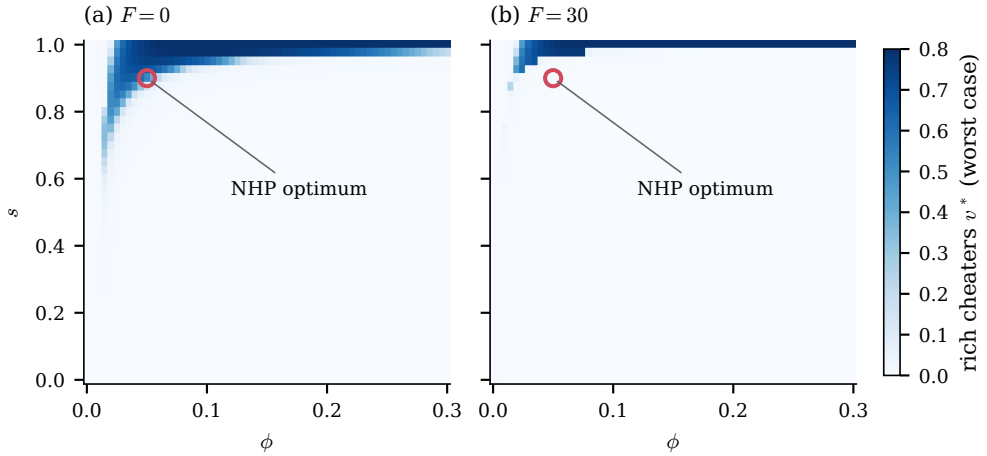

Fig. 3 Stationary fraction of rich cheaters (worst case over two initial conditions) when the rich can evade; taxes and fines proportional to income, $\tau_0^r = 0.001$, $\rho = 1$. The circle marks the NHP optimum.

Because the followers do not anticipate, the dynamic problem has no incentive to deviate later from the announced path: the restriction of an optimal open-loop policy to $[t, \infty)$ is optimal from $x(t)$ (Bellman's principle), so the open-loop solution is time-consistent. This differs from open-loop Stackelberg solutions of differential games with rational followers, which are generally time-inconsistent (Başar and Olsder 1999). Since the dynamics is deterministic and x_0 known, the optimal open-loop policy and the optimal feedback policy generate the same trajectory; they would differ only with noise or uncertainty about the state.

4.1 Static leadership with a budget cap

For a cap $\bar{\phi}$ the leader chooses $\pi = (\phi, s, F)$ with $\phi \le \bar{\phi}$ and F fixed, and $W(x) = w$ is stationary wealth. When two equilibria coexist, V^+ in (5) evaluates the compliant one and V^- the cheating one. Table 4 gives the smallest budget at which each can secure a wealthy society ($w^* > 5$). Sanctions lower it (by half with $F = 100$), and near the threshold the best split leans towards police ($s^* = 0.26$–0.44), unlike the education-heavy split optimal for large budgets. With congestion the *price of robustness* is large: a robust leader needs about twice the budget of an optimistic one, and sanctions hardly reduce it (Fig. 4).

4.2 Dynamic leadership: the optimal policy path

Problem.

Starting from the collapsed (cheating) equilibrium, the leader chooses an open-loop policy $\pi(t) = (\phi(t), s(t), F(t))$, with $\phi \in [0, 0.1]$, $s \in [0, 1]$, $F \in [0, 100]$, to maximize normalized discounted welfare

$$J[\pi] = \delta \int_0^\infty e^{-\delta t} \big[w(t) - c_F \, F(t) \, d(t) \, y(t) \big] \, dt \quad \text{s.t.} \quad \delta \int_0^\infty e^{-\delta t} \phi(t) \, dt \le \bar{\phi}, \tag{6}$$

Table 4 Minimum budget for a wealthy stationary state ($w^* > 5$) under robust and optimistic leadership.

Model	robust	optimistic	s^* at the robust threshold
NHP ($F = 0$)	0.013	0.012	0.56
$F = 30$	0.009	0.008	0.26
$F = 100$	0.006	0.005	0.44
NHP, congestion $\kappa = 0.2$	0.023	0.015	0.82
$F = 30$, congestion $\kappa = 0.2$	0.022	0.011	0.84

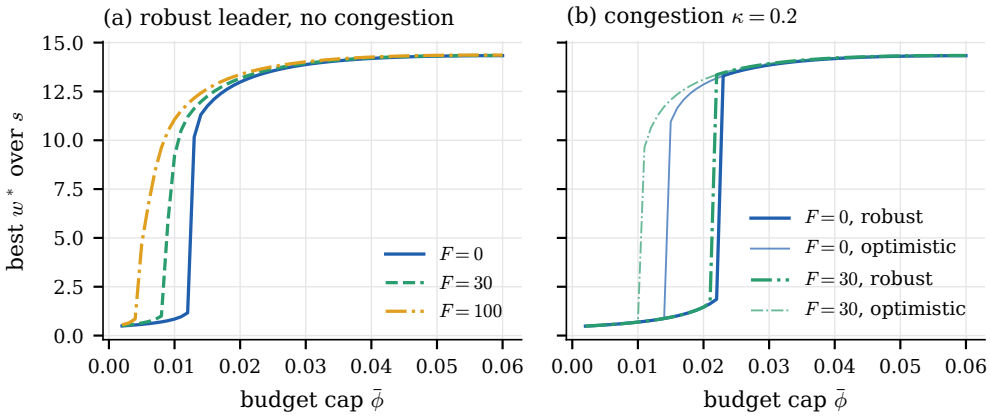

Fig. 4 Best stationary wealth for a budget cap ϕ, maximizing over s. (a) Robust leader without congestion. (b) Robust versus optimistic leader with congestion $\kappa = 0.2$.

subject to $\dot{x} = f(x, \pi)$, $x(0) = x_0$. The term $c_F F dy$ is the deadweight cost of sanctions actually imposed ($c_F = 0$: fines are transfers; $c_F = 1$: as costly as their size, e.g. prison). The present-value constraint gives the leader the same discounted effort as a constant policy $\phi \equiv \bar{\phi}$, free to distribute in time, and forces the long-run policy to be paid for.

Discretization.

Controls are piecewise constant on N intervals covering $[0, 300)$ years, and the values on the last interval are kept forever, so every admissible policy is defined on $[0, \infty)$. The admissible set is the box $[0, 0.1] \times [0, 1] \times [0, 100]$ on each interval together with the linear budget constraint, which in discrete form reads $\sum_k \omega_k \phi_k \leq \bar{\phi}$ with $\omega_k = e^{-\delta t_k} - e^{-\delta t_{k+1}}$ (and $\omega_N = e^{-\delta t_{N-1}}$ for the stationary tail). We use $N = 30$ (10-year intervals) and $N = 60$ (5-year intervals). The integral is truncated when $e^{-\delta t} < 10^{-5}$.

Method.

The problem is non-convex: from a collapsed state the gradient of J is almost flat and local methods stall in different local optima. We proceeded in three steps. (1) An exhaustive search over a structured family, "no spending until t_0; pulse (ϕ_p, s_p) for τ

years; maintenance (ϕ_m, s_m) forever", with $t_0 \in \{0, 10, \ldots, 100\}$, $\tau \in \{10, 20, 30\}$, six pulse levels, three pulse splits and five maintenance splits, and ϕ_m fixed by the budget (about 1500 policies per scenario, plus all constant policies). (2) Unrestricted SLSQP on the $N = 30$ grid (60–90 variables), started from the best structured policy and from two "immediate strike" guesses; the best local optimum is kept. (3) SLSQP on the refined $N = 60$ grid, started from the best $N = 30$ solution. Table 5 reports the three values. Unrestricted optimization improves on the structured family by 3–12%, and refining the grid adds a further 0.6–1.8%. Different starts can end in different local optima (for $\delta = 0.01$ the two starts give $J = 4.245$ and 4.053, and with sanctions 3.294 and 3.149), so we do not claim global optimality. What is robust is the *structure*: every local optimum we found, from every start, has the same three phases.

Pontryagin check.

For the $N = 60$ solutions we integrated the adjoint system of the present-value Hamiltonian $H = \delta e^{-\delta t}\big[w - c_F F dy - \nu\phi\big] + \lambda^\top f(x, \pi)$ backwards from $\lambda(T_{\text{end}}) = 0$ and computed the switching function $\sigma_\phi = \partial H/\partial\phi + \nu\delta e^{-\delta t}$, with the budget multiplier ν estimated on the intervals where $0 < \phi < 0.1$. The maximum principle requires $\phi = 0$ where $\sigma_\phi < \nu\delta e^{-\delta t}$, $\phi = 0.1$ where it is larger, and equality on singular arcs. In all four scenarios the interval averages satisfy these conditions: on interior intervals $\sigma_\phi/(\nu\delta e^{-\delta t})$ equals one within 0.9%; on the saving intervals it is at least 13% below one; on the intervals at $\phi = 0.1$ it is above one; and $|\partial H/\partial s|$ vanishes on interior values of s to within 0.7% of the scale $\nu \max \phi$ (Fig. 5c). The numerical optima are therefore extremals of the continuous problem with a bang–bang–singular structure: $\phi = 0$, then ϕ at or near its maximum, then a singular arc that is the maintenance policy.

Result: save, strike, sustain.

The numerically identified optimal policy has three phases (Fig. 5, Table 5). *Save:* nothing is spent for 40–90 years while the society is collapsed; spending there barely changes welfare, and postponing a perpetual maintenance cost frees present-value budget for a higher maintenance later. The more patient the leader, the longer the saving phase. *Strike:* a concentrated pulse crosses the saddle in 20–30 years, starting with police and moving towards education. *Sustain:* maintenance is permanent, education-heavy ($s_m \approx 0.8$–0.86) and at or above the upper fold, so hysteresis matters as a floor rather than as a saving. Impatient leaders save less and maintain at a higher level, because they value the compliant years sooner. With sanctions (F free in $[0, 100]$, $c_F = 1$) the optimizer keeps F at its cap at all times (see below); the pulse is then smaller ($\phi \le 0.046$) and maintenance cheaper ($\phi_m \approx 0.0096$). The value of the policy is 1.8–3.7 times that of the best constant policy. In the sanctions scenario the optimizer stops spending in the last interval of the grid (after 295 years), which lets the society collapse in the far future; this end effect changes J by less than 0.5%.

Sanctions and their cap.

Because $F = 100$ is reached at all times, we repeated the sanctions scenario ($N = 30$ grid, same start) with $F_{\max} = 150$ and 200. The optimum stays at the cap, $F(t) = F_{\max}$, and welfare keeps rising, $J = 3.29$, 3.84 and 4.09 for $F_{\max} = 100$, 150 and 200.

Table 5 Infinite-horizon optimal policies from a collapsed start ($\bar{\phi} = 0.012$ for $F = 0$; $\bar{\phi} = 0.008$ with sanctions, $c_F = 1$). J_{str}: best structured policy; J^*_{30}, J^*_{60}: best SLSQP optimum on 10- and 5-year grids; J_{const}: best constant policy. Phases from the $N = 60$ solution: saving length t_0, strike interval and peak budget, maintenance (ϕ_m, s_m), long-run cheater fraction.

Scenario	J_{str}	J^*_{30}	J^*_{60}	J_{const}	t_0 (yr)	strike (yr)	peak ϕ	(ϕ_m, s_m)	y_∞
$F = 0$, $\delta = 0.01$	4.15	4.25	4.27	1.15	90	90–120	0.10	(0.0147, 0.78)	0.073
$F = 0$, $\delta = 0.02$	2.88	2.94	2.96	1.15	60	65–90	0.10	(0.0161, 0.81)	0.064
$F = 0$, $\delta = 0.04$	2.00	2.07	2.09	1.15	40	40–65	0.10	(0.0188, 0.86)	0.050
$F = 100$, $\delta = 0.02$	3.00	3.29	3.35	1.21	45	45–70	0.046	(0.0096, 0.82)	0.036

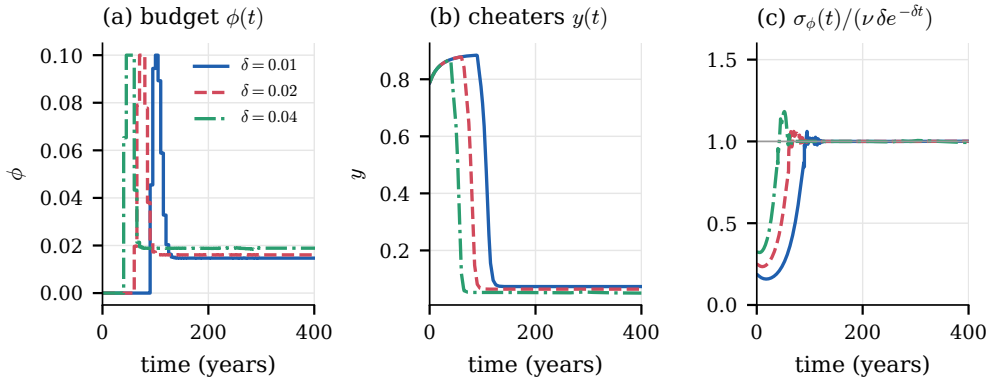

Fig. 5 Numerically optimal policies on the 5-year grid for three discount rates ($F = 0$, $\bar{\phi} = 0.012$, collapsed start). (c) Normalized switching function of the budget: below one the maximum principle prescribes $\phi = 0$, above one $\phi = 0.1$, and equal to one on the singular (maintenance) arc.

This is Becker's corner solution (Becker 1968): with a deadweight cost linear in the sanctions actually imposed, $c_F F dy$, and few cheaters in the compliant state, a higher sanction costs little and deters more. The level of F is therefore set by the cap, and the cap must be read as an institutional constraint that the model does not price: proportionality of punishment, the offenders' ability to pay, and the cost of punishing innocent people by mistake, all of which grow faster than linearly with F. To check this, we replaced the linear cost by a convex one, $c_Q F^2 dy$ with $c_Q = 0.01$ (equal to the linear cost at $F = 100$), and raised the cap to $F_{\max} = 400$. The optimal sanction is then interior in all but one interval: $F \approx 120$–270, highest during the strike and the first decades of maintenance and slowly declining as compliance consolidates, with $J = 3.90$. In every case the budget path keeps the save–strike–sustain structure (40 years of saving, a pulse with $\phi \leq 0.037$, maintenance $\phi_m \approx 0.009$–0.012): the cap on sanctions changes the value of the policy, not its timing.

Robustness.

(i) *Parameters.* With $\bar{\phi}$ placed in the middle of each parameter set's band, the save–strike–sustain structure appears in every case of Table B1 (Appendix B), with gains of 2.0–5.3 times over the best constant policy (structured search only, so these gains

are lower bounds). (ii) *Budget in money.* If the constraint limits the present value of money ϕw instead of the share ϕ (budget equal to the money spent by the share-based optimum, $\pm 25\%$), the save and strike phases persist, shorter (20–60 years) and police-heavy ($s_p = 0.3$), and welfare rises to $J = 3.25$; but whether compliance is sustained depends on the budget level, because maintaining a share of a large wealth is expensive in money. (iii) *Finite horizon.* With a 300-year horizon and an average-share budget the optimum is "strike, maintain inside the band, harvest": spending is cut in the last century, an artefact that the infinite-horizon formulation removes. (iv) *Congestion.* With $\kappa = 0.2$ a collapsed society does not escape within 300 years for $\bar{\phi} \leq 0.030$, and for $\bar{\phi} = 0.036$ only after two centuries: the budget is a share of a small wealth, so even $\phi = 0.1$ buys little police. This is a poverty trap that only resources not proportional to w could break.

5 Spatial structure

A cellular-automaton version of the NHP model was introduced by Meacci (2015); here we add payoff-driven imitation and local budgets. In a large society, taxes are collected and spent locally and imitation is local.

Construction.

Agents occupy the cells of an $L_y \times L_x$ square lattice with periodic boundaries; each cell is a poor taxpayer (P), a rich taxpayer (R), a cheater (C) or, in the extension of Section 3.3, a rich cheater. Each cell x also carries a local budget $w(x)$, which represents taxes collected and spent in its neighbourhood and obeys

$$\dot{w}(x) = \bar{z}(x) + a\,\bar{u}(x) - (\theta + \phi)\,w(x), \tag{7}$$

where $\bar{z}, \bar{u}$ are the fractions of P and R cells in the $(2R+1)\times(2R+1)$ box centred at x. All rates are evaluated with the local budget: $\alpha = \alpha_0 + \alpha_1 s\phi w(x)$, $d = k(1-s)\phi w(x)$ and $\Delta\pi = g - Fd$. In a time step Δt every cell draws one uniform random number and changes state with probability rate$\times\Delta t$:

$$\begin{aligned} &\mathrm{P} \to \mathrm{R}: \ \alpha, && \mathrm{P} \to \mathrm{C}: \ \tau_0 + \tau_1\, c(x)\,[\Delta\pi]_+, \\ &\mathrm{R} \to \mathrm{P}: \ \beta, && \mathrm{C} \to \mathrm{P}: \ \mu + d + \tau_1\, p(x)\,[-\Delta\pi]_+, \end{aligned}$$

where $c(x)$ and $p(x)$ are the fractions of cheaters and of poor taxpayers among the 8 Moore neighbours and $[\cdot]_+ = \max(\cdot, 0)$. The two exits from P are mutually exclusive: a P cell with random number u becomes R if $u < \alpha\Delta t$, becomes C if $\alpha\Delta t \leq u < (\alpha + r_{PC})\Delta t$, where r_{PC} is the P→C rate, and stays P otherwise. This is pairwise proportional imitation on the lattice: a reviser looks at a random neighbour and copies it with probability proportional to the payoff advantage. Since $[\Delta\pi]_+ - [-\Delta\pi]_+ = \Delta\pi$, replacing c and p by global fractions recovers the mean-field flow $\tau_1 \Delta\pi\, yz$ of (3). All cells are updated synchronously and (7) is advanced by an explicit Euler step of the same Δt. We use $\Delta t = 0.1$ yr (0.05 yr with sanctions); for the policies simulated all transition probabilities per step are below 0.03, so the discrete-time

process approximates the continuous-time one. Homogeneous 100×100 lattices started at a mean-field equilibrium outside the band stay within 0.02 in y of it over 1500 years.

The Maxwell point.

On a 300×30 strip whose left half starts at the compliant equilibrium and right half at the cheating one (cells drawn independently with the equilibrium class fractions, w at its equilibrium value), the two interfaces move at constant speed, and their direction depends only on the policy (Fig. 6a). The speed vanishes at a *Maxwell point* ϕ_M (Pomeau 1986) inside the band: above it compliance invades, below it cheating does. A large spatial society is therefore governed by one threshold instead of two. To quantify the statistical uncertainty we ran 6 independent realizations (different initial lattices and random streams) at 5 values of ϕ around ϕ_M for each case, 390 runs in total. Front speeds vary between realizations with a standard deviation of 0.001–0.02 cells per year, but ϕ_M, obtained by interpolating the mean speed and bootstrapping over realizations, is determined to within a 95% interval of 3–11% of the band width (Table 6). For $R = 3$, ϕ_M lies in the lower 37% of the band (between about 3% and 37% of its width above the lower fold) for all s and for $F = 0$ and $F = 30$ (Fig. 6b); with congestion ($\kappa = 0.2$, $s = 0.5$), $\phi_M = 0.0183$ in the band $[0.0147, 0.0254]$. Escape from collapse above ϕ_M happens by spontaneous nucleation of compliant domains, and its time is independent of lattice size for $L \geq 80$; local spending pulses barely accelerate it.

A deterministic front theory.

The Maxwell point can be predicted without simulating the automaton. Replacing each cell by its expected densities gives a lattice dynamical system: every site carries (z, y, w) and obeys (1)–(3), except that the imitation term uses the Moore average of y and the wealth equation uses the box averages of z and u. For slowly varying fields the Moore average is $y + \frac{3}{8}h^2\Delta y$, so in the continuum limit the cheater density obeys a reaction–diffusion equation, nonlinear through the factor z of the imitation flow, coupled to a wealth field with a nonlocal (box) source. A planar front between the two stable states then travels at a speed $c(\phi)$ that changes sign at a unique ϕ_M in the band. Because the system is not of gradient type, ϕ_M is not given by an equal-area rule, but it is computed by integrating the front (Appendix A). Table 6 compares the theory with the automaton in three groups. (i) *Policies, $R = 3$.* Where the automaton gives a Maxwell point ($F = 0$ with $s \geq 0.5$, and all four cases with $F = 30$), theory and automaton differ by 0–11% of the band width (at most 2.2×10^{-4} in ϕ). For $F = 0$, $s = 0.3$, the narrowest band (8×10^{-4} wide), compliance already invades at the lowest ϕ sampled, so the automaton only gives the bound $\phi_M \leq 0.0125$, about 2×10^{-4} below the theoretical 0.0127. (ii) *Congestion.* With $\kappa = 0.2$ the difference is 2% of the band. (iii) *Radius of the local budget.* The theoretical value is unchanged when the grid is refined ($h = 1$, 0.5, 0.25) and hardly depends on R, whereas the automaton does. For $R = 1$ the local budget averages only nine cells, and its fluctuations make compliance so fragile that cheating fronts still advance at the upper mean-field fold (speed -0.057 ± 0.005 cells per year): ϕ_M lies above the band, and the comparison with a mean-field band is not meaningful. For $R = 2$ the automaton lies 16% of the band

Table 6 Maxwell points: cellular automaton (CA; $R = 3$ unless stated; 6 realizations per ϕ, 95% bootstrap interval) versus the deterministic front theory. Last column: difference as a share of the bistable band. For $F = 0$, $s = 0.3$ compliance already invades at the lowest ϕ sampled; for $R = 1$ cheating still invades at the upper fold.

Case	band $[\phi_{\mathrm{lo}}, \phi_{\mathrm{hi}}]$	ϕ_M (CA)	95% interval	ϕ_M (theory)	diff.
$F = 0$, $s = 0.3$	$[0.0125, 0.0133]$	≤ 0.0125	—	0.0127	26%
$F = 0$, $s = 0.5$	$[0.0111, 0.0129]$	0.0115	$[0.0115, 0.0116]$	0.0115	0%
$F = 0$, $s = 0.7$	$[0.0114, 0.0137]$	0.0119	$[0.0118, 0.0120]$	0.0119	1%
$F = 0$, $s = 0.9$	$[0.0149, 0.0170]$	0.0153	$[0.0152, 0.0154]$	0.0155	8%
$F = 30$, $s = 0.3$	$[0.0076, 0.0091]$	0.0082	$[0.0081, 0.0082]$	0.0080	11%
$F = 30$, $s = 0.5$	$[0.0076, 0.0099]$	0.0083	$[0.0083, 0.0084]$	0.0081	9%
$F = 30$, $s = 0.7$	$[0.0085, 0.0114]$	0.0093	$[0.0092, 0.0094]$	0.0091	5%
$F = 30$, $s = 0.9$	$[0.0125, 0.0155]$	0.0133	$[0.0132, 0.0134]$	0.0132	3%
$s = 0.5$, $R = 1$	$[0.0111, 0.0129]$	> 0.0129	—	0.0116	—
$s = 0.5$, $R = 2$	$[0.0111, 0.0129]$	0.0118	$[0.0117, 0.0119]$	0.0116	16%
$s = 0.5$, $R = 5$	$[0.0111, 0.0129]$	0.0112	$[0.0112, 0.0113]$	0.0115	17%
$s = 0.5$, $R = 8$	$[0.0111, 0.0129]$	0.0112	$[0.0112, 0.0113]$	0.0115	21%
$s = 0.5$, $\kappa = 0.2$	$[0.0147, 0.0254]$	0.0183	$[0.0180, 0.0185]$	0.0180	2%

above the theory. For $R = 5$ and $R = 8$ it lies 17–21% of the band below it (3–4×10^{-4} in ϕ), well outside the statistical uncertainty: with large local budgets the automaton is slightly more favourable to compliance than the deterministic theory, presumably because random compliant clusters are not destroyed by fluctuations of the budget and help fronts advance. The deterministic theory is therefore an approximation of the automaton, accurate to a few 10^{-4} in ϕ for $R \geq 2$, and not its exact large-R limit. Noise also makes fronts in the automaton several times faster than the deterministic ones (Fig. 6a), but it hardly moves the point where they stop.

Locality matters.

In the automaton ϕ_M decreases with the radius of the local budget, from above the upper mean-field fold ($R = 1$) to 0.0118 ($R = 2$) and 0.0112 ($R \geq 5$): with very small tax bases fluctuations make compliance fragile. Rich evasion forms clusters and is roughly halved relative to the mean field (at the NHP optimum, $v \approx 0.29$ instead of 0.53), but without sanctions the education-heavy policy still loses about a third of its wealth.

6 Discussion

Treating the NHP population model as an evolutionary game changes three of its answers. First, *how much to invest* has a sharp minimum, the upper fold for a robust government or the Maxwell point in a spatial society, and sanctions lower it by the factor in Proposition 2. Second, *how to split* the budget depends on the size of the budget and on who can cheat: education-heavy splits are optimal for large budgets only if the rich cannot evade or sanctions are in place, while near the threshold police-heavy splits are better. Third, *when to spend* is itself an instrument: hysteresis makes concentrated spending more effective than the same budget spread evenly.

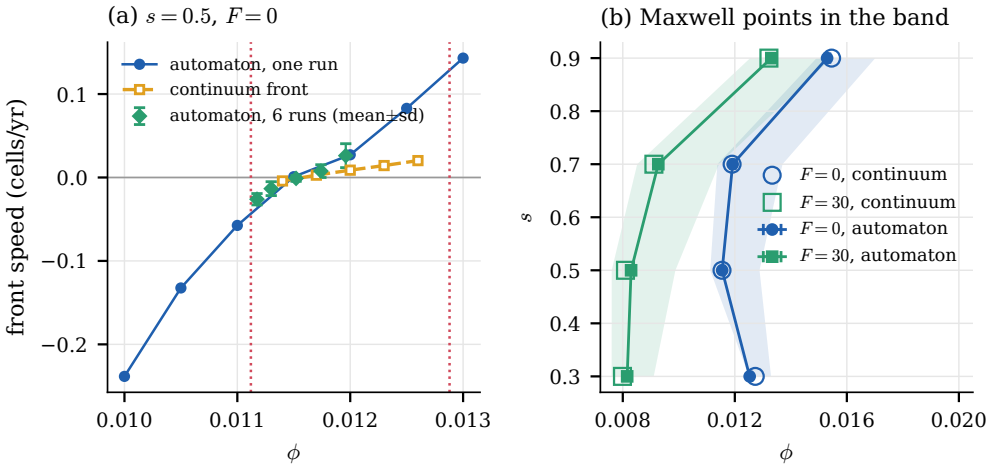

Fig. 6 (a) Front speed versus ϕ ($s = 0.5$, $F = 0$): one automaton realization over a wide range, mean ± standard deviation of 6 realizations near the Maxwell point, and the deterministic front theory; dotted lines: mean-field folds. (b) Maxwell points versus s for $F = 0$ and $F = 30$ (filled with 95% intervals: automaton; open: front theory) over the mean-field bistable bands (shaded).

The save–strike–sustain structure is related to the history dependence of optimal policies in non-convex control problems with multiple steady states (Skiba 1978; Grass et al. 2008), and to optimal dynamic law enforcement (Feichtinger et al. 2002). The new ingredient here is that the state is an evolutionary population of imitators, so the leader steers basins of attraction rather than a single steady state. The saving phase depends on measuring the budget as a share of wealth; in money terms it shortens but persists.

Limitations.

Two assumptions deserve a closer look. *Sanctions are not revenue.* Adding the collected fines Fdy to the wealth equation can only help compliance, because the extra revenue is largest exactly when cheaters are being detected and finances more police. We checked this at $s = 0.5$: the bistable band moves from $[0.0077, 0.0098]$ to $[0.0074, 0.0089]$ for $F = 30$ and from $[0.0045, 0.0064]$ to $[0.0043, 0.0051]$ for $F = 100$. The upper fold, the budget needed to escape a collapsed state, falls most, so our estimates of the minimum budgets and of the benefit of sanctions are conservative. A full treatment would bound fines by the cheaters' income. *Open-loop policy.* As explained in Section 4, in a deterministic model with a known initial state open-loop and feedback optimal policies coincide along the optimal path, and the open-loop solution is time-consistent. The open-loop assumption matters when the state is observed with error, parameters are uncertain or the dynamics is noisy (as in finite populations, Appendix B): a policy timed to cross a saddle may then cross too early or too late, and a feedback rule ("strike when y falls below a threshold") would be more robust. The saving phase is the part most exposed to this, since its length is set by the discount rate and not by the state. Other limitations are that imitation across classes is not modelled, that the leader cannot borrow from outside the society, the natural way

to break the congestion trap, and that the spatial leader problem, where to spend and not only how much, remains open.

Appendix A Numerical methods and validation

Equilibria and folds.

Zeros of G (Lemma 1) are bracketed on a grid of 2500–4000 points in $(0, 1)$ and refined with Brent's method; stability follows from the eigenvalues of a finite-difference Jacobian. Fold curves are obtained, for each s, by scanning ϕ and bisecting (22 steps) the boundaries of the set where two stable equilibria exist.

Optimal control.

Fixed-step RK4 ($\Delta t = 0.1$ yr) vectorized over a batch of policies; the integral is truncated when $e^{-\delta t} < 10^{-5}$. Structured policies are evaluated exhaustively. SLSQP uses forward finite-difference gradients (step 10^{-5}, all components in one batched integration), the exact linear budget constraint, box bounds, tolerance 10^{-10} and at most 300 iterations; all reported runs terminated successfully. *Pontryagin check:* the state is recomputed with RK4 ($\Delta t = 0.05$ yr); the adjoint $\dot{\lambda} = -\partial H/\partial x$ is integrated backwards with Heun's method from $\lambda(T_{\text{end}}) = 0$; Jacobians and $\partial H/\partial \pi$ are computed by complex-step differentiation. Interval averages of σ_ϕ are ratios of integrals over each control interval (for the stationary tail, over $[t_{N-1}, T_{\text{end}}]$).

Cellular automaton.

Rules as in Section 5; periodic boundaries in both directions, synchronous updates, $\Delta t = 0.1$ yr (0.05 yr with sanctions). *Fronts:* $L_y \times L_x = 30 \times 300$ (checked with $L_y = 15$ and 60, which change ϕ_M by less than 4×10^{-5}); half compliant, half cheating, so the periodic strip contains two interfaces; the position is the number of columns with cheater fraction below 0.45, and the speed is half the slope of a linear fit over $t \in [150, 1000]$ years, using only times at which both phases occupy more than five columns. *Replicates:* 6 realizations at 5 values of ϕ spanning $\pm 25\%$ of the band around a first estimate of ϕ_M (clipped to the band); ϕ_M is the zero of the linear interpolation of the mean speed, and its 95% interval comes from 2000 bootstrap resamples of realizations. *Nucleation:* $L \times L$ lattices with $L = 40, 80, 160$ started in the cheating state, 3 realizations per size; the escape time is the first time at which half the lattice is compliant. *Homogeneous checks:* 100×100 lattices, 1500 years.

Front theory.

The deterministic lattice system is integrated with explicit Euler ($\Delta t = 0.05$ yr) on a periodic one-dimensional domain of length 300 with two fronts, mesh $h = 1$ (checked with $h = 0.5$ and 0.25), box averages with fractional end weights; the speed is fitted from the length of the compliant region over $t \in [0, 1500]$, and ϕ_M is found by nine bisection steps inside the band.

Table A1 Validation against NHP.

Quantity	NHP	This work
y^*, w^* at $\phi = 0$	0.89, 0.45	0.888, 0.449
$w_{\max}$ without cheaters ($s = 1$) and its ϕ	14.5, 0.05	14.50, 0.050
y^* at $s = 1$ (independent of k)	0.87	0.872
Bistability threshold in (k, s) at $\phi = 0.01$	$k_c \approx 1.24$	$k_c \approx 1.25$

Appendix B Sensitivity and finite populations

Bistability requires enough imitation.

Bistability disappears when the imitation strength is low: for $\tau_1 = 0.05$ no s is bistable, for $\tau_1 = 0.06$ only $s \in [0.65, 0.85]$, and the bistable range widens with τ_1 ($[0.25, 0.95]$ for $\tau_1 = 0.09$). Results that rely on hysteresis therefore apply to societies where cheating spreads by imitation at least at this rate.

Table B1 Sensitivity of the bistable band at $s = 0.5$ and of the infinite-horizon optimal policy ($\delta = 0.02$, $\bar{\phi}$ at the middle of the band; rows with $F = 30$ use $c_F = 1$). Gain: best structured policy over the best constant policy.

Case	band ($s = 0.5$)	J^*	gain	t_0, τ (yr)	(ϕ_m, s_m)
NHP	$[0.0111, 0.0129]$	2.88	2.5	60, 30	(0.0151, 0.8)
$\tau_1 \times 2$	$[0.0147, 0.0242]$	2.43	5.3	60, 30	(0.0354, 0.7)
$k = 0.5$	$[0.0150, 0.0180]$	2.74	2.4	60, 30	(0.0177, 0.7)
$k = 2$	$[0.0081, 0.0087]$	2.87	2.0	60, 30	(0.0096, 0.8)
$\tau_0 = 0.001$	$[0.0090, 0.0114]$	2.62	2.3	80, 20	(0.0261, 0.9)
$F = 30$, $\lambda = \tau_1$	$[0.0076, 0.0099]$	2.39	3.4	60, 30	(0.0118, 0.8)
$F = 30$, $\lambda = 0.05$	$[0.0090, 0.0112]$	2.46	3.1	80, 20	(0.0255, 0.9)
$F = 30$, $\lambda = 0.2$	$[0.0058, 0.0080]$	2.33	3.4	80, 10	(0.0197, 0.8)

Finite populations.

Simulations with N individuals (tau-leaping, 40 replicates, 20 000 years) at the middle of the band ($s = 0.5$) show that bistability survives: for $N \geq 1000$ no replicate crosses the saddle in either direction. For $N = 200$ (NHP model), all 40 replicates started in the cheating state cross the saddle towards compliance within 20 000 years (mean first-passage time 3700 years), whereas only 5 of the 40 replicates started in the compliant state (12%) cross towards cheating. Noise-induced transitions are therefore asymmetric: small communities are more often rescued than ruined by noise.

Acknowledgements. This work is dedicated to the memory of Mario Primicerio (1940–2025), whose work on the mathematical modelling of crime and cheating inspired it.

Declarations

- Funding: No funding was received for conducting this study.
- Competing interests: The authors declare no competing interests.
- Data and code availability: The data and the Python code used to obtain the results are available from the corresponding author on reasonable request.
- Author contributions: All authors contributed equally to the conception, analysis, numerical work and writing of this study, and read and approved the final manuscript.

References

Allingham MG, Sandmo A (1972) Income tax evasion: a theoretical analysis. Journal of Public Economics 1(3–4):323–338

Andrei AL, Comer K, Koehler M (2014) An agent-based model of network effects on tax compliance and evasion. Journal of Economic Psychology 40:119–133. https://doi.org/10.1016/j.joep.2013.01.002

Andreozzi L (2004) Rewarding policemen increases crime. Another surprising result from the inspection game. Public Choice 121(1):69–82

Antoci A, Russu P, Zarri L (2014) Tax evasion in a behaviorally heterogeneous society: an evolutionary analysis. Economic Modelling 42:106–115. https://doi.org/10.1016/j.econmod.2014.06.002

Avenhaus R, von Stengel B, Zamir S (2002) Inspection games. In: Aumann RJ, Hart S (eds) Handbook of Game Theory with Economic Applications, vol 3. Elsevier, Amsterdam, p 1947–1987

Başar T, Olsder GJ (1999) Dynamic Noncooperative Game Theory, Classics in Applied Mathematics, vol 23, 2nd edn. SIAM, Philadelphia

Becker GS (1968) Crime and punishment: an economic approach. Journal of Political Economy 76(2):169–217

Berestycki H, Nadal JP (2010) Self-organised critical hot spots of criminal activity. European Journal of Applied Mathematics 21(4–5):371–399

Bloomquist KM (2006) A comparison of agent-based models of income tax evasion. Social Science Computer Review 24(4):411–425

Brandt H, Hauert C, Sigmund K (2006) Punishing and abstaining for public goods. Proceedings of the National Academy of Sciences USA 103(2):495–497

Bruni C, Nuño JC, Primicerio M (2013) What if criminals optimize their choice? Optimal strategies to defend public goods. Physica A 392(4):840–850. https://doi.

org/10.1016/j.physa.2012.10.002

Feichtinger G, Grienauer W, Tragler G (2002) Optimal dynamic law enforcement. European Journal of Operational Research 141(1):58–69

Grass D, Caulkins JP, Feichtinger G, et al (2008) Optimal Control of Nonlinear Processes: With Applications in Drugs, Corruption, and Terror. Springer, Berlin

Han TA, Tran-Thanh L (2018) Cost-effective external interference for promoting the evolution of cooperation. Scientific Reports 8:15997. https://doi.org/10.1038/s41598-018-34435-2

Hauert C, Traulsen A, Brandt H, et al (2007) Via freedom to coercion: the emergence of costly punishment. Science 316(5833):1905–1907

Hofbauer J, Sigmund K (1998) Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge

Hokamp S, Pickhardt M (2010) Income tax evasion in a society of heterogeneous agents – evidence from an agent-based model. International Economic Journal 24(4):541–553. https://doi.org/10.1080/10168737.2010.525994

Kleshnina M, Streipert S, Brown JS, et al (2023) Game theory for managing evolving systems: challenges and opportunities of including vector-valued strategies and life-history traits. Dynamic Games and Applications 13(4):1130–1155. https://doi.org/10.1007/s13235-023-00544-5

Korobow A, Johnson C, Axtell R (2007) An agent-based model of tax compliance with social networks. National Tax Journal 60(3):589–610

Loridan P, Morgan J (1996) Weak via strong Stackelberg problem: new results. Journal of Global Optimization 8(3):263–287. https://doi.org/10.1007/BF00121269

Meacci L (2015) How and how much to invest for fighting cheaters: from an ODE to a cellular automata model. Mathematical Sciences 9(3):137–143. https://doi.org/10.1007/s40096-015-0160-y

Nuño JC, Herrero MA, Primicerio M (2008) A triangle model of criminality. Physica A 387(12):2926–2936

Nuño JC, Herrero MÁ, Primicerio M (2010) Fighting cheaters: how and how much to invest. European Journal of Applied Mathematics 21(4–5):459–478. https://doi.org/10.1017/S0956792510000094

Nuño JC, Herrero MA, Primicerio M (2011) A mathematical model of a criminal-prone society. Discrete and Continuous Dynamical Systems – Series S 4(1):193–207. https://doi.org/10.3934/dcdss.2011.4.193

Perc M, Jordan JJ, Rand DG, et al (2017) Statistical physics of human cooperation. Physics Reports 687:1–51

Pickhardt M, Prinz A (2014) Behavioral dynamics of tax evasion – a survey. Journal of Economic Psychology 40:1–19

Pickhardt M, Seibold G (2014) Income tax evasion dynamics: evidence from an agent-based econophysics model. Journal of Economic Psychology 40:147–160

Pomeau Y (1986) Front motion, metastability and subcritical bifurcations in hydrodynamics. Physica D 23(1–3):3–11

Sandholm WH (2010) Population Games and Evolutionary Dynamics. MIT Press, Cambridge, MA

Schlag KH (1998) Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. Journal of Economic Theory 78(1):130–156

Short MB, D'Orsogna MR, Pasour VB, et al (2008) A statistical model of criminal behavior. Mathematical Models and Methods in Applied Sciences 18(suppl.):1249–1267

Sigmund K, De Silva H, Traulsen A, et al (2010) Social learning promotes institutions for governing the commons. Nature 466(7308):861–863

Skiba AK (1978) Optimal growth with a convex-concave production function. Econometrica 46(3):527–539

Staňková K, Brown JS, Dalton WS, et al (2019) Optimizing cancer treatment using game theory: a review. JAMA Oncology 5(1):96–103. https://doi.org/10.1001/jamaoncol.2018.3395

Stein A, Salvioli M, Garjani H, et al (2023) Stackelberg evolutionary game theory: how to manage evolving systems. Philosophical Transactions of the Royal Society B 378(1876):20210495. https://doi.org/10.1098/rstb.2021.0495

Stuckey K, Dua R, Ma Y, et al (2022) Optimal dynamic incentive scheduling for Hawk–Dove evolutionary games. Physical Review E 105(1):014412. https://doi.org/10.1103/PhysRevE.105.014412

Szabó G, Fáth G (2007) Evolutionary games on graphs. Physics Reports 446(4–6):97–216

Szolnoki A, Szabó G, Perc M (2011) Phase diagrams for the spatial public goods game with pool punishment. Physical Review E 83(3):036101

Tsebelis G (1989) The abuse of probability in political analysis: the Robinson Crusoe fallacy. American Political Science Review 83(1):77–91

Zaklan G, Westerhoff F, Stauffer D (2009) Analysing tax evasion dynamics via the Ising model. Journal of Economic Interaction and Coordination 4(1):1–14. https://doi.org/10.1007/s11403-008-0043-5