Disarmament as a Signal: Firearms Regulation when Citizens Cannot Verify the Government's Type

Philip Kroos*

August 13, 2026

Abstract

A government that restricts civilian firearms to reduce crime and a government that disarms citizens ahead of coercion take the same observable action. I study what follows from that single fact. In a signalling model with a privately informed government, restriction cannot separate types over a wide region of the parameter space: the predatory type always has a mimicking incentive, and the informational content of firearms policy is therefore zero in equilibrium. Citizens respond to restriction by arming more, not less, which partially offsets the policy. The welfare-relevant object is the ex ante choice between constitutional entrenchment and political discretion. Entrenchment costs a known amount under a liberal government and insures against strategic disarmament under a predatory one, so it is optimal above a threshold prior π^* that the calibration puts at 0.034 but that ranges from 0.017 to 0.19 across plausible parameters. The result is not that arms deter states: the model never requires civilians to win a fight. Two findings run against the motivating hypothesis. Credible universal disarmament remains first best. And the value of entrenchment is hump-shaped in institutional quality, negative in both consolidated democracies and failed states, because the probability that a constitution is needed and the probability that it binds move in opposite directions.

Keywords: signalling, commitment, constitutional design, deterrence, firearms **JEL:** D72, D82, K42, H11

*philip.kroos@web.de. Replication code at github.com/Philip-Kroos/arms-commitment. All errors are my own.

1 Introduction

Governments regulate the private ownership of weapons, and the standard economic case for doing so is straightforward. Firearms in private hands generate externalities that individual owners do not internalise, so private armament exceeds the social optimum and restriction raises welfare.

This paper does not dispute that argument. It asks what happens to it when citizens cannot verify why the restriction is being imposed.

The observation driving the paper is that two very different governments take the same action. A government maximising citizen welfare restricts firearms because private armament is excessive. A government contemplating coercion restricts firearms because a disarmed population is cheaper to coerce. Both issue the same regulation. Citizens observing the regulation cannot tell which motive produced it.

This turns firearms policy into a signalling problem, and the signalling problem has consequences that the standard analysis does not capture. If restriction raises the posterior that the government is predatory, it raises the private incentive to arm out of fear, which works against the restriction itself. And if the predatory type always has an incentive to mimic the liberal type, then no restriction can ever credibly convey benign intent, so a government that genuinely wants to reduce gun violence has no way to say so.

What the paper does not argue. The familiar version of the constitutional argument for civilian arms is that an armed population deters state coercion by being able to resist it. That argument is weak, and I do not make it. A modern state's coercive capacity dominates any plausible distribution of civilian small arms; the collective action problem in organised resistance is severe; and the historical record of armed civilian resistance to consolidated states is not encouraging. The mechanism here does not require civilians to win, or to fight, or even to be capable of fighting effectively. It requires only that disarming them is *cheaper than not disarming them* from the perspective of a government contemplating coercion, and that citizens know this.

Model and results. Section 2 sets out a game with a continuum of citizens and a government of privately known type, liberal or predatory, with prior π on the latter. Citizens' armament decisions respond both to the mechanical cost of regulation and to the posterior it induces. Under a calibration in which private armament exceeds the social optimum — the condition without which no government would restrict and the model would have nothing to say — five results follow.

First, the predatory type strictly prefers to restrict even with no welfare motive whatever, and restriction moves armament across the threshold at which coercion becomes profitable. This is strategic disarmament, and it is derived rather than assumed.

Second, in the baseline region restriction pools: there is no equilibrium in which the two types choose different policies. Firearms regulation carries no information about the government's type. This is conditional, not universal: when the restriction technology is weak relative to the private protective value of arms, a separating equilibrium exists, and Figure 3 maps the boundary.

Third, the informational effect can overturn the mechanical one. When arms provide sufficient private protection against coercion, a suspicious citizenry arms hard enough to restore deterrence despite the restriction, at $\delta \geq 0.31$ in the calibration. Restriction is then self-defeating in the strict sense that it fails to reduce armament.

Fourth, the ex ante comparison between constitutional entrenchment and political discretion has a threshold structure. Entrenchment costs 0.031 in welfare under a liberal government and is worth 0.882 under a predatory one, a ratio of 28 to 1, which puts the critical prior at $\pi^* = 0.034$. That number should not be taken seriously as a point estimate; it is driven almost entirely by the ratio of the harm from state coercion to the harm from gun violence, which the model does not identify and which is a normative input. It ranges from 0.017 to 0.19 across the parameters in Table 4.

Fifth, and least expected, the value of entrenchment is hump-shaped in institutional quality. Better institutions make a predatory government less likely, which reduces the need for a constitutional constraint, and simultaneously make a constitution more likely to be obeyed, which increases its value. Neither force is assumed to dominate. Computing the product gives a maximum at intermediate institutional quality and negative values at both ends (Figure 1). Entrenchment is worth nothing in a consolidated democracy, because it is never needed, and nothing in a failed state, because it is never honoured.

Two results against the motivating hypothesis. The project began from the hypothesis that civilian armament generates deterrence benefits that standard policy evaluation misses. Two findings cut against it and are reported as such. Credible universal disarmament is first best in the model ($W_A = 0.023$ against $W_F = 0.014$ for constitutional protection): if compliance could be made universal, including by offenders, restriction would dominate. The case for entrenchment is strictly second best. And the hump-shape implies that the countries where the constitutional argument is usually made — stable, wealthy, institutionally strong democracies — are precisely the ones where the model says the argument is weakest.

Relation to the literature. Three strands are relevant, and the paper's claim to novelty rests on the gap between them rather than on any one of them.

The economics of crime and self-defence has modelled the citizen–offender interaction as a game with private information about armament, generating deterrence externalities, target substitution and multiple equilibria [Marceau, 1997, Anonymous, 2015]. That literature is the source of the crime-side parameters d and h in Section 2, which I take as an input and do not

claim to advance.

The political economy of repression models the interaction between an elite and a citizenry with a revolution constraint, and establishes commitment problems as the central mechanism in institutional change [Acemoglu and Robinson, 2000, 2006]. Acemoglu and Robinson [2000] in particular allow the elite's type to be private information. The present model is a direct descendant. Its difference is the object of inference: in Acemoglu and Robinson [2000] citizens are uncertain about the elite's *capacity* to repress, whereas here they are uncertain about its *intent*, and the signal they observe is a policy that a benign government has an independent and legitimate reason to adopt. It is the existence of that legitimate reason that destroys the signal.

The empirical literature on disarmament as a coercive instrument establishes that governments do disarm populations for strategic reasons and that it works [Zhukov, 2016]. That paper supplies the mechanism's empirical plausibility and its identification strategy is the closest thing to a template for testing the vertical channel here.

What the paper adds is the interaction. Once a benign motive for restriction exists, the predatory type can hide behind it, and the informational consequence — not the military one — is what makes constitutional entrenchment valuable. This reframing matters because it survives the objection that sinks the standard argument: it does not require civilians to be able to resist.

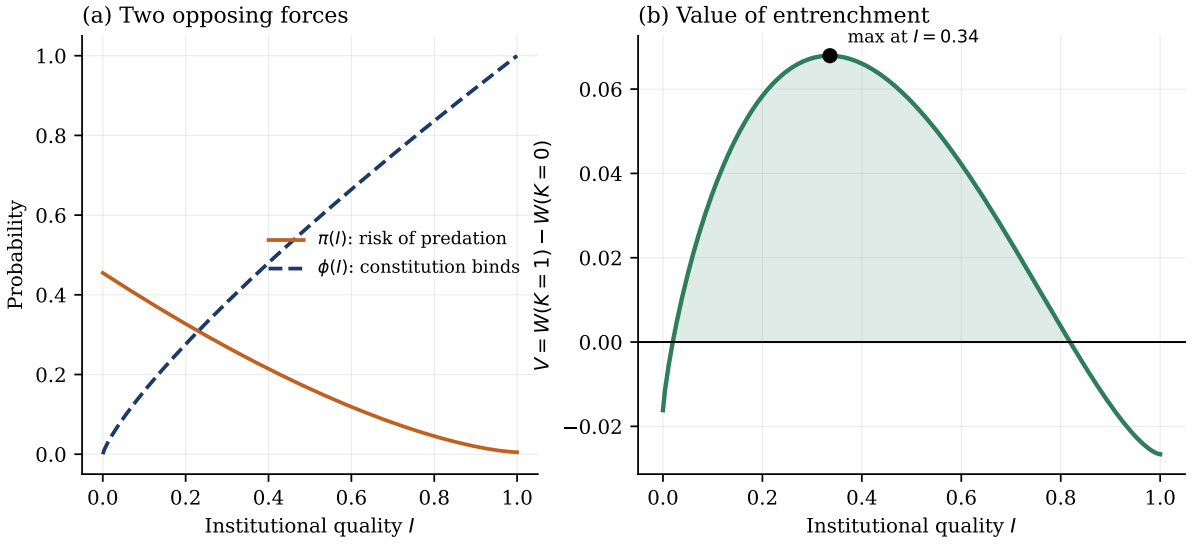

Figure 1: **The result that runs against the premise.** Better institutions make a predatory government less likely (π falls) and a constitution more likely to be obeyed (ϕ rises). Neither force is assumed to dominate. Their product peaks at intermediate institutional quality and is negative at both ends: entrenchment is worth nothing in a consolidated democracy, where it is never needed, and nothing in a failed state, where it is never honoured.

Section 2 presents the model, Section 3 the results, Section 4 an empirical design that is a design and not a set of findings, Section 5 the limitations, which are substantial.

2 Model

2.1 Environment

A unit continuum of citizens indexed by i and a government G. The government's type $\theta \in \{L, P\}$ is drawn with $\Pr(\theta = P) = \pi$ and privately observed. The liberal type L maximises citizen welfare. The predatory type P obtains gross benefit B from coercion.

Timing.

$t = 0$ Behind the veil, before θ is drawn, a constitutional rule $K \in \{0, 1\}$ is chosen. $K = 1$ removes firearms restriction from the government's action set.

$t = 1$ θ is realised and privately observed.

$t = 2$ G chooses $R \in \{0, 1\}$. If $K = 1$, $R = 0$ is imposed with probability ϕ and the constraint fails with probability $1 - \phi$.

$t = 3$ Citizens observe R, form posterior $\mu = \Pr(\theta = P \mid R)$ and choose armament.

$t = 4$ G chooses whether to coerce.

The parameter ϕ deserves a sentence. A constitution is a document, and a government willing to coerce its citizens is not obviously deterred by one. Setting $\phi < 1$ is the honest modelling choice: constitutional constraints bind types that are already partly constrained. It also turns out to drive the main comparative-static result.

2.2 Citizens

Citizen i has private cost $c_i \sim U[0, \bar{c}]$ of acquiring a firearm, scaled by $(1 + \tau R)$ under restriction. Restriction raises the cost for compliant citizens only; non-compliant actors are unaffected, which is the asymmetry emphasised in the horizontal literature and which enters here through the deterrence parameter d.

Citizen i arms if

$$\underbrace{b_H}_{\text{crime protection}} + \underbrace{\mu\,\sigma\,(D_u - D_a)}_{\text{coercion protection}} \;\geq\; c_i\,(1 + \tau R), \tag{1}$$

where σ is the perceived probability of coercion and $D_u > D_a$ are the losses from coercion when unarmed and armed. Aggregate armament is

$$A(R, \mu, \sigma) = \frac{b_H + \mu\sigma(D_u - D_a)}{\bar{c}\,(1 + \tau R)}. \tag{2}$$

Equation (2) contains the paper's mechanism. Restriction enters twice and with opposite signs: it raises the denominator, and by raising μ it raises the numerator.

Note that the second term in (1) is a *private* benefit. No individual's firearm changes the probability of coercion; the deterrence effect is a pure public good and is therefore under-provided relative to the citizens' collective interest. This is not incidental. It is why a constitutional rule, which operates on aggregate armament directly, can improve on the decentralised outcome even when citizens are free to arm.

2.3 Government

The cost of coercion is

$$C(A) = \gamma + \lambda A, \tag{3}$$

increasing in civilian armament with $\lambda > 0$. The predatory type coerces if $B > C(A)$, giving the deterrence threshold

$$A^* = \frac{B - \gamma}{\lambda}. \tag{4}$$

The liberal type never coerces and maximises $w(A) = dA - \frac{1}{2}hA^2$, the net social value of armament in the crime game, maximised at $A_L = d/h$.

Assumption 1 (Over-armament)**.** $b_H/\bar{c} > d/h$: unregulated private armament exceeds the social optimum.

Assumption 1 is not innocuous and is worth flagging, because violating it destroys the model rather than merely changing its conclusions. If citizens under-arm relative to the social optimum, no welfarist government restricts, there is nothing for a predatory type to hide behind, and the signalling problem does not arise. My first calibration violated it, which inverted the entire mechanism before I noticed. The assumption is also the standard premise of the case *for* gun control, so adopting it concedes the strongest form of the opposing argument.

2.4 Equilibrium

The solution concept is perfect Bayesian equilibrium. Off-path beliefs are set to $\mu = 1$: an unexpected restriction is read as predatory. This is the belief least favourable to sustaining restriction, so any pooling-on-restriction result is not an artefact of a convenient refinement.

Armament and the coercion probability are jointly determined and are solved as a fixed point of (2) and the coercion rule. Where no pure-strategy fixed point exists the predatory type mixes at exactly the indifference point; that region is where citizens' fear is partly self-confirming and is reported rather than assumed away.

3 Results

All results below are verified numerically in `src/verify.py`; the calibration is Table 1. Under it the social optimum is $A_L = 0.300$, unregulated armament is 0.550, armament under restriction is 0.289, and the deterrence threshold is $A^* = 0.450$.

Proposition 1 (Strategic disarmament)**.** *The predatory type strictly prefers restriction even with no welfare motive, and restriction moves armament across the deterrence threshold:* $A(R{=}0) = 0.550 > A^* = 0.450 > A(R{=}1) = 0.312$.

Restriction is not merely correlated with subsequent coercion; in the model it is what makes coercion profitable. The predatory type's payoff from restricting is 0.097 against 0.000 from not restricting, and the difference is entirely the change in the cost of coercion.

Proposition 2 (Pooling)**.** *In the baseline region the unique equilibrium is pooling on restriction:* $(R_L, R_P) = (1, 1)$. *No separating equilibrium exists. Firearms policy carries no information about the government's type.*

The liberal type cannot signal benign intent by restricting, because the predatory type does the same thing; and cannot signal it by not restricting, because that sacrifices the welfare gain that motivated the policy. There is no message available. This is the paper's central claim and it is the reason the argument does not depend on civilians being able to fight.

Proposition 2 is conditional. Figure 3 maps the regions: when τ is small relative to $\delta = D_u - D_a$, a separating equilibrium exists. The conditionality is stated rather than suppressed because the failure region is informative — it says that where arms are highly protective and regulation weakly effective, gun policy *does* reveal intent.

Proposition 3 (Self-defeating restriction)**.** *For* $\delta \geq 0.31$ *in the calibration, a maximally suspicious citizenry arms sufficiently under restriction to restore deterrence:* $A(R{=}1, \mu{=}1) \geq A^*$.

Proposition 4 (Deterrence paradox)**.** *Across the parameter range, higher armament lowers both the probability of coercion and observed armed resistance. Comparing high- and low-armament regions:* $A = 0.503$ *versus* 0.204, $\Pr(\textit{coerce}) = 0.000$ *versus* 1.000, *observed resistance* 0.000 *versus* 0.204.

Proposition 4 is an obstacle rather than a finding. The mechanism is most effective exactly where the data show nothing, so the absence of armed resistance is uninformative about the deterrence value of arms. Any empirical design must confront this, and Section 4 argues that the observable margin is government behaviour, not citizen behaviour.

Proposition 5 (Constitutional threshold)**.** $V \equiv W(K{=}1) - W(K{=}0)$ *crosses zero once in* π, *negative below. Under the calibration* $\pi^* = 0.034$, *decomposing into a peacetime cost of* 0.031 *and an insurance value of* 0.882.

The threshold is low because the ratio of the two components is 28 to 1, and that ratio is a direct consequence of scaling the harm from state coercion ($D_u = 1$) large relative to the welfare loss from over-armament (≈ 0.03). *The model does not identify this ratio.* It is the single most important free parameter and it is essentially a normative judgement about how bad state coercion is relative to gun violence. Table 4 shows π^* moving from 0.017 to 0.189 over plausible values, and for $h \leq 0.6$ or $\lambda \geq 1.1$ entrenchment dominates for every π. A reader who thinks D_u should be an order of magnitude smaller gets a correspondingly higher threshold, and nothing in the analysis prevents that.

Proposition 6 (Institutional hump)**.** *Let π be decreasing and ϕ increasing in institutional quality I. Then $V(I)$ attains an interior maximum, at $I = 0.335$ with $V = +0.068$ under the parameterisation of Figure 1, and is negative at both endpoints: $V(0) = -0.016$, $V(1) = -0.027$.*

This is the result most worth testing and the least congenial to the argument that usually motivates the question. Constitutional protection of civilian arms is worth least in consolidated democracies, where a predatory government is improbable, and least in weak states, where a constitution does not bind. Its value is concentrated in the middle band — societies with institutions strong enough to be respected but not strong enough to be relied on.

Proposition 7 (Second best)**.** *Credible universal disarmament is first best: $W_A = 0.023 > W_F = 0.014$. No feasible restriction policy attains both the peacetime optimum and deterrence. Among feasible regimes, constitutional protection dominates disarmament under weak institutions by* 0.375*.*

The ordering in Table 5 is $W_A > W_D > W_E = W_F > W_B > W_C$. If compliance could be made universal, including among offenders, restriction would dominate and this paper's argument would not apply. The case for entrenchment is a second-best argument conditional on incomplete compliance and uncertain institutions, and it should not be reported as anything stronger.

4 Empirical strategy

This section is a design. No estimates are reported and none should be inferred.

The identification problem. Proposition 4 rules out the natural test. Armed resistance is rare precisely where deterrence works, so its absence is uninformative. The observable margin is not citizen behaviour but *government sequencing*: does civilian disarmament systematically precede coercive episodes, conditional on the factors that predict both?

Design 1: sequencing (preferred). Construct an event history of civilian firearm restrictions and coercive episodes at the country–year level. Test whether restriction predicts subsequent repression after conditioning on regime type, prior violence and state capacity. **Threat:** restriction

and repression share unobserved drivers, most obviously an anticipated threat to public order, which is exactly what a liberal government would also respond to. This is a severe confound and the design does not solve it. **Data:** V-Dem for institutional and repression measures, Small Arms Survey for civilian prevalence [ACCESS TO VERIFY], GunPolicy.org for legislative dates [ACCESS TO VERIFY].

Design 2: within-country implementation variation. Following Zhukov [2016], exploit sub-national variation in the intensity of disarmament campaigns. Stronger because it holds the national regime fixed; weaker because it speaks to counterinsurgency rather than constitutional commitment.

Design 3: horizontal channel only. The citizen–offender predictions are testable with US state-level variation in carry regimes. This is well-trodden ground with a contested literature, and the honest position is that it would replicate rather than extend.

Falsification. The model is falsified if: offenders do not respond to the probability that a target is armed; restriction does not raise the subjective probability of future coercion in survey data; disarmament shows no sequencing relationship with coercion after controls; or the estimated ratio of coercion harm to firearm harm falls below the level that makes π^* exceed any plausible prior. The last is the most likely to bind.

5 Limitations

Four, in descending order of severity.

The key ratio is not identified. Everything quantitative rests on the harm from state coercion relative to the harm from gun violence. The model takes this as a parameter. It is a normative input, and a reader who sets it differently gets a different answer.

The vertical channel may not be testable. Proposition 4 implies the mechanism hides where it works, and the sequencing design in Section 4 has a confound it cannot remove.

Collective action is assumed away. Aggregate armament enters the cost of coercion directly. In reality, dispersed individuals face a severe coordination problem in converting private arms into organised resistance. Introducing it would reduce λ, and Table 4 shows the results are sensitive to λ.

Entrenchment may have costs the model omits. A constitutional right to arms may itself raise the probability of political violence, or harden political conflict in ways that damage the institutions it is meant to protect. The model has no channel for this and would need one before its policy implications were taken seriously.

6 Conclusion

Firearms regulation is a signal as well as a policy, and it is a signal that cannot be made credible, because the government with the worst intentions has the strongest incentive to send it. That is the paper's contribution, and it is an information-economics contribution rather than a claim about the military value of small arms.

What follows is narrower than the argument usually made. Universal disarmament remains first best, at 0.023 against 0.014 for constitutional protection. Entrenchment is a second-best response to incomplete compliance and uncertain government type, valuable above a threshold prior of 0.034 in the baseline that ranges to 0.189 and that the model cannot pin down, and valuable mainly in societies whose institutions are neither strong enough to make it unnecessary nor weak enough to make it futile.

The mechanism is not specific to firearms. Any privately held technology that raises the cost of coercion and that a government might restrict for legitimate reasons — strong encryption is the obvious contemporary case — generates the same inference problem and the same argument for removing the decision from ordinary political discretion. Whether that generalisation survives contact with those settings is the natural next question.

References

Daron Acemoglu and James A. Robinson. Democratization or repression? *European Economic Review*, 44(4–6):683–693, 2000.

Daron Acemoglu and James A. Robinson. *Economic Origins of Dictatorship and Democracy*. Cambridge University Press, Cambridge, 2006.

Anonymous. The deterrence effects of gun laws in games with asymmetric skills and information. *Review of Law and Economics*, 11(3):435–452, 2015. Full author list to verify against the published article.

Nicolas Marceau. Self-selection and violence in the market for crime. *International Review of Law and Economics*, 17(2):193–201, 1997.

Yuri M. Zhukov. Taking away the guns: Forcible disarmament and rebellion. *Journal of Peace Research*, 53(2):242–258, 2016.

Tables and Figures

Table 1: Calibration. The binding requirement is Assumption 1: unregulated armament (0.550) must exceed the social optimum (0.300)

	Value
d	0.3000
h	1.0000
b_H	0.5500
$\bar{c}$	1.0000
τ	0.9000
π	0.1500
B	0.5150
γ	0.2000
λ	0.7000
D_u	1.0000
D_a	0.7200
ϕ	1.0000

Table 2: Signalling stage. Only pooling on restriction survives: firearms policy carries no information about the government's type

	Type	Equilibrium	u_L	u_L dev	u_P	u_P dev
$(R_L, R_P) = (0, 0)$	pooling	no	0.0137	0.0356	-0.0000	0.0092
$(R_L, R_P) = (0, 1)$	separating	no	0.0137	0.0356	0.0092	-0.0000
$(R_L, R_P) = (1, 0)$	separating	no	0.0449	0.0137	-0.0000	0.1124
$(R_L, R_P) = (1, 1)$	pooling	yes	0.0449	0.0137	0.0969	-0.0000

Off-path beliefs set to $\mu = 1$, the value least favourable to sustaining restriction.

Table 3: Welfare decomposition. The 28:1 ratio of insurance value to peacetime cost is what drives π^* down to 0.034 — and it is a calibration choice

	baseline
$W(K{=}0)$	-0.0920
$W(K{=}1)$	0.0137
V	0.1057
Peacetime cost	0.0312
Insurance value	0.8816
Ratio	28.2710
π^*	0.0340

Table 4: Comparative statics. π^* ranges from 0.017 to 0.189; for a small gun externality or strong deterrence technology, entrenchment dominates for every π

parameter	value	π^*	Note
Harm from coercion D_u	0.3	0.113	
	0.5	0.068	
	1.0	0.034	
	2.0	0.017	
Gun externality h	0.6		always entrench
	1.0	0.034	
	1.6	0.103	
	2.4	0.189	
Constitution binds ϕ	0.3	0.105	
	0.6	0.056	
	0.85	0.040	
	1.0	0.034	
Restriction effectiveness τ	0.5	0.032	
	0.9	0.034	
	1.5	0.030	
	2.5	0.022	
Deterrence technology λ	0.4	0.876	
	0.7	0.034	
	1.1		always entrench
	1.6		always entrench

"Always entrench" means $V > 0$ for every $\pi \in (0, 1)$.

Table 5: Regime comparison. Credible universal disarmament is first best (0.023 against 0.014): the case for entrenchment is strictly second best

	W
A: credible universal disarmament	0.0233
B: civilian disarmament, armed offenders	-0.1699
C: disarmament, weak institutions	-0.3617
D: broad armament, discretion	0.0200
E: strong institutions, armament	0.0137
F: constitutional protection	0.0137

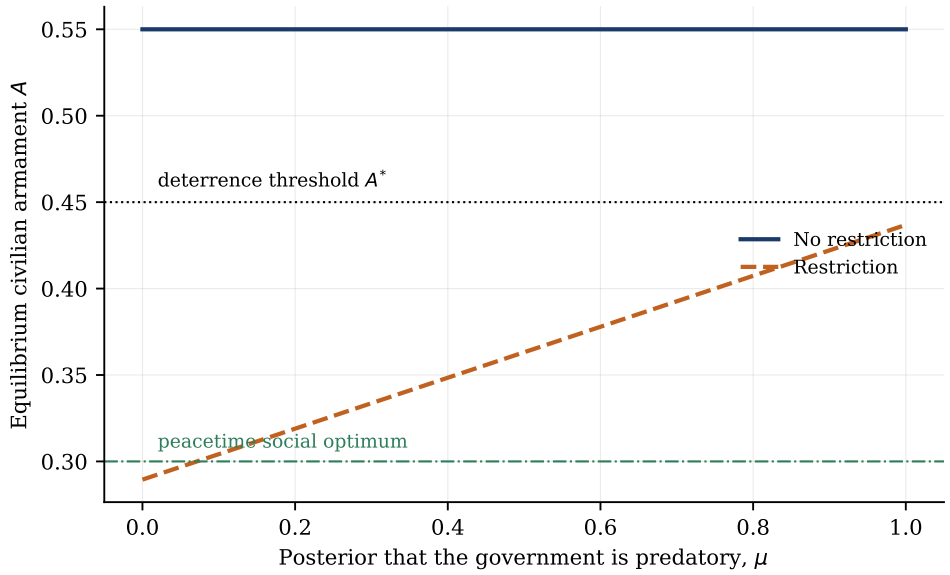

Figure 2: Restriction shifts armament down mechanically and up informationally.

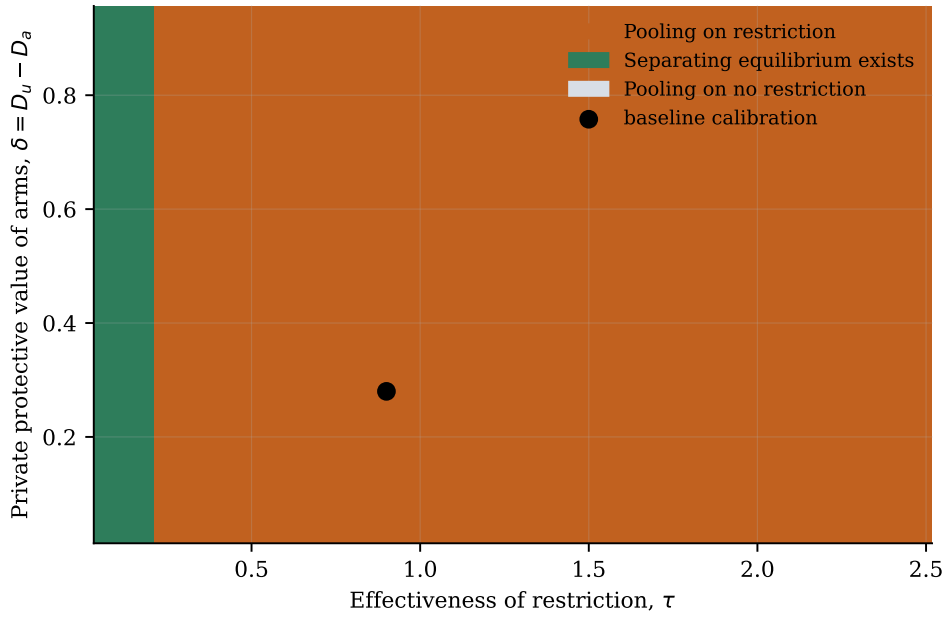

Figure 3: Where restriction pools and where it separates. Pooling is the baseline case but not universal.

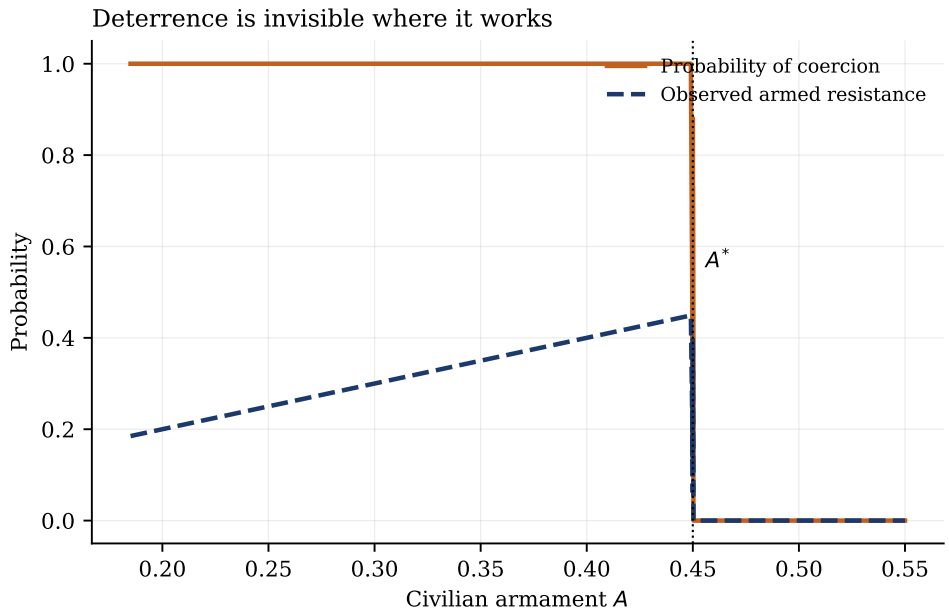

Figure 4: The mechanism is invisible where it works.

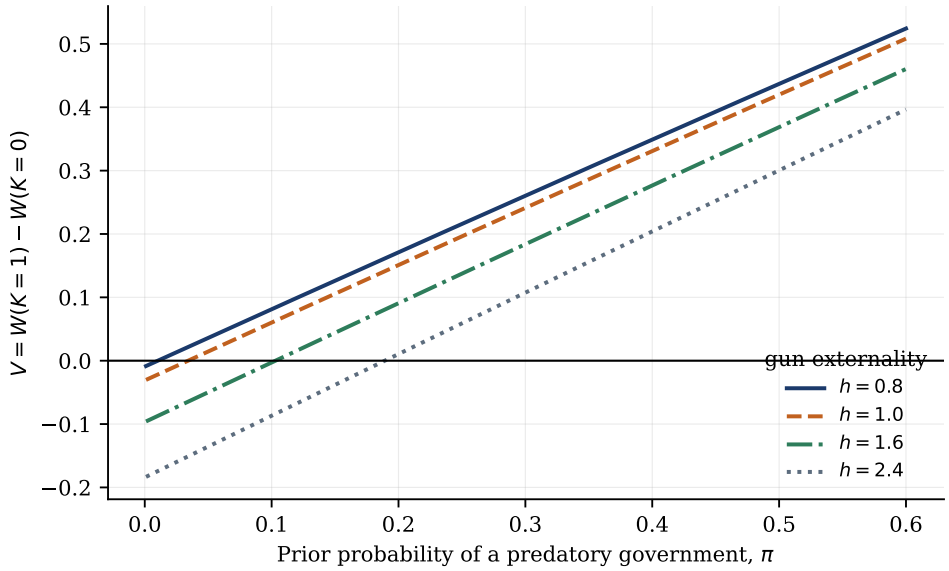

Figure 5: Value of entrenchment against the prior, for four values of the gun externality.

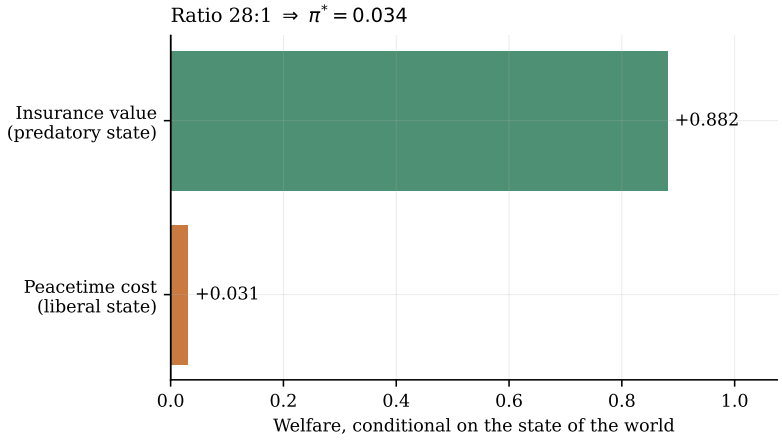

Figure 6: Where the low threshold comes from. The 28:1 ratio is a calibration choice, not a finding.