How Jeremy Bentham would defend against coordinated attacks

Ole Jann* and Christoph Schottmüller**

*Nuffield College Oxford
**University of Copenhagen, Tilec

Outline

Big picture

How to exercise power?

How to maintain order?

What do we look at?

Game theoretic model of

- 1 central player ("warden")
- threat of coordinated attack by N "prisoners"

- warden
  - how much *costly ressources* ("guard level") to fight off possible attack?
  - what *information* about guard level to release in order to exploit prisoner's coordination problem? (prison design)

What about Bentham? I

Jeremy Bentham (1748-1832)

What about Bentham? II

- Bentham's suggestion: Panopticon
  - no information on guard level
  - keep prisoners separate (to hamper coordination)
- Bentham's claims
  - coordination to breakout will never be achieved
  - regardless of how many/whether guard(s) are on duty "[. . .] *so far from it, that a greater multitude than ever were yet lodged in one house might be inspected by a single person*"
  - can be applied to everything: schools, factories, hospitals. . .

Relation to literature in social sciences and game theory

- Foucault: enforcement by panopticon allowed “accumulation of men” necessary for industrial take off
- add endogenous information structure to *global games* (Carlsson and van Damme 1993, Morris and Shin. . .). typical applications:
  - central bank defending currency peg against speculators (Morris and Shin 1998)
  - government defending against coup d’état (Chassang and i Miquel 2009)

Main result

- Bentham was right if the number of prisoners is high
  - secrecy of guard level optimally exploits coordination problem
  - in equilibrium warden uses minimal guard level
  - probability of breakout is almost zero nevertheless

Main result

- Bentham was right if the number of prisoners is high
  - secrecy of guard level optimally exploits coordination problem
  - in equilibrium warden uses minimal guard level
  - probability of breakout is almost zero nevertheless

- rough intuition
  - "matching pennies" incentives
  - transparency vs. secrecy
    - self-confirming vs. self-defeating beliefs
  - law of large numbers: quite precise idea of how many prisoners revolt
    - suppose many
    - employ more guards
    - no one wants to revolt... contradiction

Model

- one warden
  - sets a guard level $\gamma \in \Re_+$
  - payoff:
    - $-B-\gamma$ if there is a break out
    - $-\gamma$ if there is no break out
- N prisoners
  - actions: "revolt" (r), "not revolt" (n)
  - payoff:

	break out	no break out
r	$b > 0$	$-q < 0$
n	0	0

- breakout iff strictly more than γ prisoners revolt
- Assumption: $B \geq N + 1$
  (prevent breakout under complete info)

Solution concept

- Nash equilibrium (in mixed strategies)
  - warden chooses probability distribution over guard levels
  - prisoners simultaneously choose probability p of revolting
  - choice of each player maximizes his expected payoff (taking other players' choices as given)

Information

		Guard level observable	
		Yes	No
Coordination problem	No	(1a) Benchmark	(1b) Benchmark
	Yes	(2) Transparency	(3) Panopticon

Table: The four information structures we consider.

Benchmarks (perfect prisoner coordination)

- guard level observed
  - all revolt if $\gamma < N$
  - none revolts otherwise
  - equilibrium: $\gamma = N$
- guard level unobserved
  - either all or none revolt
  - γ either 0 or N
  - mixed strategy equilibrium
- equilibrium payoffs
  - warden: $-N$
  - prisoner: 0

Panopticon (guard level unobserved, no coordination) I

Lemma

Equilibria are prisoner symmetric, i.e. all prisoners revolt with the same probability.

- probability p to revolt
- number revolting prisoners: binomial distribution
- only mixed strategy equilibria

Panopticon (guard level unobserved, no coordination) II

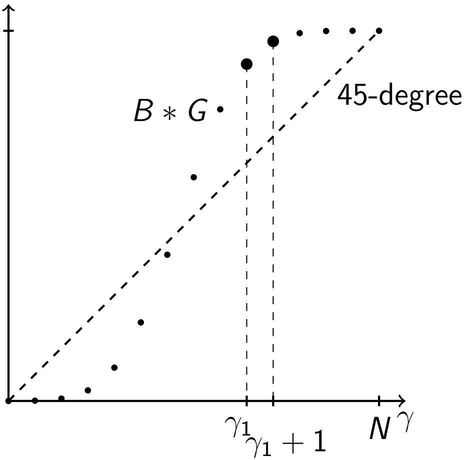
Lemma

In equilibrium, the warden mixes between two adjacent guard levels γ_1 and γ_1+1 where $\gamma_1 \in \{0, \ldots, N-1\}$.

- possibly multiple equilibria

Panopticon (guard level unobserved, no coordination) III

- warden payoff: $-(1-G(\gamma))B-\gamma$ (binomial distrib. is G)



Main Result

Theorem (Bentham was right)

Take b and q as given. Let N be sufficiently large and B such that assumption 1 holds.
Then, the warden mixes between 0 and 1 in the unique equilibrium of the panopticon.
The probability of a breakout is arbitrarily close to zero for sufficiently high N.

Main Result (rough intuition) I

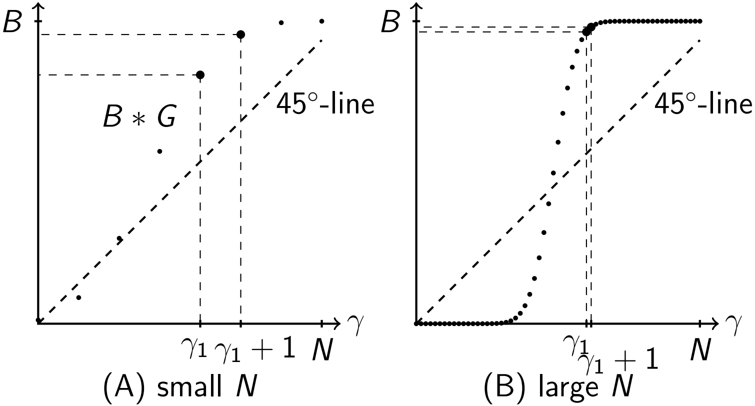


Figure: An illustration of the main result.

- warden indifference: $Bg(\gamma + 1) = 1$

Main Result (rough intuition) II

- for high N distribution of revolting prisoners G concentrated around mode pN
- around mode marginal utility of $\gamma \uparrow$ high
- γ_1 substantially above mode
- probability that more than γ_1 prisoners revolt low
- prisoner strictly prefers not to revolt

- what is different for $\gamma_1 = 0$?

Main Result (rough intuition) II

- for high N distribution of revolting prisoners G concentrated around mode pN
- around mode marginal utility of $\gamma \uparrow$ high
- γ_1 substantially above mode
- probability that more than γ_1 prisoners revolt low
- prisoner strictly prefers not to revolt

- what is different for $\gamma_1 = 0$?
  - revolt is dominant strategy if $\gamma_1 = 0$
  - 0-1 equilibrium: less coordination game but one-to-one "matching pennies"

Comparison: Transparency I (guard level observed, no coordination)

say warden chooses guard level γ

- if $\gamma \geq N$: not revolt (dominant)
- if $\gamma < 1$: revolt (dominant)
- if $1 \leq \gamma < N$
  - either all revolt in subgame equilibrium
  - or none revolts in subgame equilibrium

Comparison: Transparency I (guard level observed, no coordination)

say warden chooses guard level γ

- if $\gamma \geq N$: not revolt (dominant)
- if $\gamma < 1$: revolt (dominant)
- if $1 \leq \gamma < N$
  - either all revolt in subgame equilibrium
  - or none revolts in subgame equilibrium

- equilibrium selection as in global games
- result (roughly):
  - play r if and only if $\gamma < \lceil bN/(q+b) \rceil$
  - warden sets $\gamma = \lceil bN/(q+b) \rceil$
- warden payoff linearly decreasing in N

Comparison: Transparency II (guard level observed, no coordination)

Result

Panopticon yields higher expected warden payoff (and welfare) for N large than transparency.

Discussion

- How to save a currency peg?
  - keep your foreign currency reserves secret!
  - what about "forward guidance" and transparency?
- Minimal enforcement
  - What about massive police presence at demonstrations/football etc.?
  - Extension: minimum guard level

Discussion

- How to save a currency peg?
  - keep your foreign currency reserves secret!
  - what about "forward guidance" and transparency?
- Minimal enforcement
  - What about massive police presence at demonstrations/football etc.?
  - Extension: minimum guard level

Corollary (Panopticon with minimal guard requirement)

Suppose the warden has to set a guard level of at least $\gamma_{min} \geq 1$ with probability of at least $z_{min} > z_{eq}$.
Then there is a unique equilibrium in the panopticon in which the warden sets γ_{min} with probability z_{min} (and $\gamma = 0$ with probability $1 - z_{min}$) and prisoners revolt with zero probability.

Robustness/Extensions

- payoff when unsuccessfully revolting might depend on guard level
  - revolutions: punishment if seen
  - say $-q - \rho\gamma/N$
  - everything goes through: behave as watched because you might be watched
- payoff of not revolting depends on whether there is a breakout
  - revolution: punishment of non revolting (everything goes through)
  - free riding: can destroy strategic complementarity (destroys results)
- some randomness in breakout probability
  - prob of breakout is $\beta \mathbb{1}_{m>\gamma} + (1-\beta)m/N$
- attackers have different sizes

Conclusion

- coordinated attack model where central player chooses
  - defense level
  - information about defense level
- how to exercise power through the choice of information structure
- optimal to keep defense level secret (for N large etc.)

Proof (sketch) I

Candidate equilibrium:

(p, γ) such that

- $Bg(\gamma + 1) = 1$ (warden indifferent between γ and $\gamma + 1$)

write $B = \alpha(N + 1)$ (recall: $\alpha \geq 1$ by assumption)

rewrite candidate eq. condition:

$$\binom{N}{\gamma + 1} p^{\gamma+1}(1 - p)^{N-\gamma-1} = \frac{1}{\alpha(N + 1)}$$

Proof (sketch) II

Lemma

The probability $1 - G_N(\gamma)$ that $\gamma + 1$ or more prisoners revolt in any equilibrium candidate converges to zero as N grows large.

Proof: Chernoff bound

$$1 - G_N(\gamma) \leq \left(\frac{N}{\gamma+1}\right)^{\gamma+1} \left(\frac{N}{N-\gamma-1}\right)^{N-\gamma-1} p^{\gamma+1}(1-p)^{N-\gamma-1}$$

for any candidate eq. this becomes:

$$1 - G_N(\gamma) \leq \left(\frac{N}{\gamma+1}\right)^{\gamma+1} \left(\frac{N}{N-\gamma-1}\right)^{N-\gamma-1} \frac{1}{\alpha(N+1)\binom{N}{\gamma+1}}$$

Proof (sketch) III

$$1 - G_N(\gamma) \leq \left(\frac{N}{\gamma+1}\right)^{\gamma+1} \left(\frac{N}{N-\gamma-1}\right)^{N-\gamma-1} \frac{1}{(N+1)\binom{N}{\gamma+1}}$$

let $m = \gamma + 1$:

$$1 - G_N(\gamma) \leq \frac{1}{\binom{N}{m}(m/N)^m((N-m)/N)^{N-m}} \frac{1}{(N+1)}$$

denominator minimized by $m = N/2$ (probability mass of a binomial distribution with $p = m/N$ evaluated at mode)

$$1 - G_N(\gamma) \leq \frac{2^N}{\binom{N}{N/2}\alpha(N+1)} \leq \frac{\sqrt{2N}}{(N+1)}$$

as $\binom{N}{N/2} \geq 2^N/\sqrt{2N}$, RHS converges to zero as $N \to \infty$ □

Transparency model (guard level observed, no coordination), details I

- warden chooses guard level with trembling hand $\gamma \sim N(\tilde{\gamma}, \varepsilon')$
- prisoner observes signal drawn from uniform distribution on $[\gamma - \varepsilon, \gamma + \varepsilon]$

Lemma

Let $\varepsilon' > 0$. Assume that $bN/(q+b) \notin \mathbb{N}$ and define

$$\theta^* = \left\lceil \frac{bN}{q+b} \right\rceil .$$

Then for any $\delta > 0$, there exists an $\bar{\varepsilon} > 0$ such that for all $\varepsilon \leq \bar{\varepsilon}$, a player receiving a signal below $\theta^* - \delta$ will play r and a player receiving a signal above $\theta^* + \delta$ will play n.

Transparency model (guard level observed, no coordination) , details II

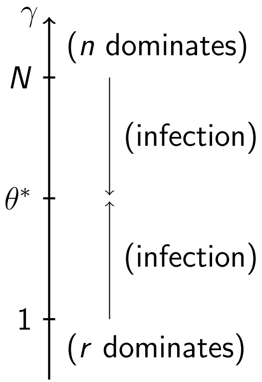


Figure: Infection of beliefs among prisoners

Other results I

Theorem (high disutility of breakout B)

Unless a single guard deters prisoners in the transparency model, the warden is better off in the panopticon if B is sufficiently large.

Other results I

Theorem (high disutility of breakout B)

Unless a single guard deters prisoners in the transparency model, the warden is better off in the panopticon if B is sufficiently large.

- only 0-1 equilibrium exists for high B
- any other γ_1:
  - for B high enough, γ_1 is only optimal if p is very low
  - prisoners strictly prefer not to revolt

Other results II

Theorem (incentives to revolt b/q)

For b/q sufficiently high, the warden payoff is $-N$ in all models.

- *Suppose $B^{\frac{N-1}{N}} > N$: Then, for $b/q \in (N-1, B^{\frac{N-1}{N}} - 1)$, the warden's payoff in every equilibrium of the panopticon model is higher than in the equilibrium of the transparency model.*
- *Suppose $N > B^{\frac{N-1}{N}}$: Then, for $b/q \in (B^{\frac{N-1}{N}} - 1, N-1)$, there exists an equilibrium in the panopticon model in which the warden's equilibrium payoff is lower than in the transparency model.*