

# How Jeremy Bentham would defend against coordinated attacks

Ole Jann\* and Christoph Schottmüller\*\*

\*University of Copenhagen → Oxford

\*\*University of Copenhagen, Tilac

# Outline

- 1 Introduction
- 2 Model
- 3 Results
- 4 Discussion
- 5 Conclusion
- 6 Appendix

# What do we look at?

- 1 central player (“warden”)
- threat of coordinated attack by  $N$  “prisoners”
  
- warden
  - how much *costly resources* (“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?

- 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. . .

## Is this (related to) economics?

- 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... )
  - 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
  - law of large number: 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  $\gamma \in \mathbb{R}_+$
  - 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  $\gamma$  prisoners revolt
- Assumption:  $B \geq N + 1$   
(prevent breakout under complete info)



# 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.

## Transparency (guard level observed, no coordination)

- 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

# Transparency (guard level observed, no coordination)

- 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  $\gamma = \lceil bN/(q+b) \rceil$

# Panopticon (guard level unobserved, no coordination) I

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

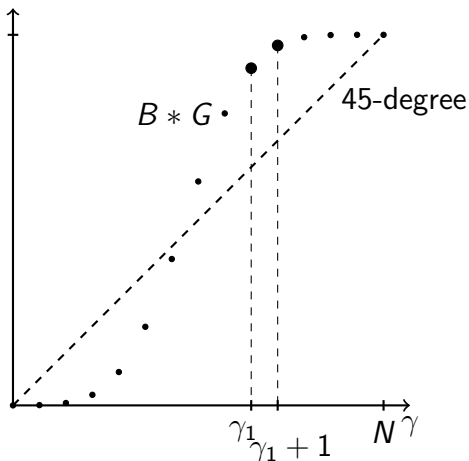
## Lemma

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

- possibly multiple equilibria

## Panopticon (guard level unobserved, no coordination) II

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



# Main Result

## Theorem (Bentham was right)

*Let  $N$  be sufficiently large. Then, the warden mixes between 0 and 1 in the unique equilibrium of the panopticon model. The warden's payoff is higher in this equilibrium than in the transparency model.*

*In the panopticon, the probability of a breakout is arbitrarily close to zero for sufficiently high  $N$ .*

## Main Result (rough intuition)

- for high  $N$  distribution of revolting prisoners  $G$  concentrated around mode  $pN$
  - around mode marginal utility of  $\gamma \uparrow$  high
  - $\gamma_1$  substantially above mode
  - probability that more than  $\gamma_1$  prisoners revolt low
  - prisoner strictly prefers not to revolt
- 
- what is different for  $\gamma_1=0$ ?

# Main Result (rough intuition)

- for high  $N$  distribution of revolting prisoners  $G$  concentrated around mode  $pN$
- around mode marginal utility of  $\gamma \uparrow$  high
- $\gamma_1$  substantially above mode
- probability that more than  $\gamma_1$  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”



# Discussion

- How to save a currency peg?
  - keep your foreign currency reserves secret!
  - what about “forward guidance” and transparency?
- Minimal enforcement
  - Bentham and Foucault
  - What about massive police presence at demonstrations/football etc.?

# 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.)

# Benchmark (no coordination problem)

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

## 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

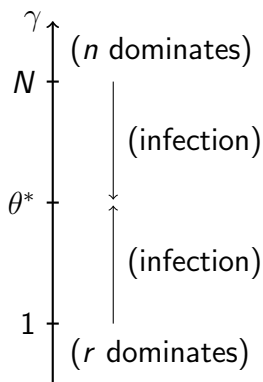


Figure: Infection of beliefs among prisoners

## 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  $\gamma_1$ :
  - for  $B$  high enough,  $\gamma_1$  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.*