Waiting for Fake News

Ralph Boleslavsky

MWET
November 2023

Background

To a cop, the explanation is ... always simple. If you got a dead body and you think the brother did it, you're gonna find out you're right.

—Verbal Kint, *The Usual Suspects*

Background

- In the US, the decision to file charges against a suspect is made by a local prosector, based on an investigation conducted by the police.
- Imagine that a long investigation has produced circumstantial evidence against suspect, but it has not produced direct evidence.
- The prosecutor and police officer agree on the interpretation of evidence, and both want to see justice done.
- When deciding whether to pursue case, prosecutor accounts for costs not borne by the police officer (opportunity cost, direct cost, reputation).
- Prosecutor hesitant to pursue the case based on inconclusive evidence. Needs to be more convinced that suspect is guilty.
- Extensive investigation uncovered no direct evidence, case closed soon

Background

- If police officer sufficiently convinced by circumstantial evidence, he or she may be tempted to *fabricate* evidence of guilt
  - Planting physical evidence
  - Coercing false confession
  - Misrepresenting or manipulating forensic analysis
  - Procuring false witness testimony or identifications
  - (all from National Registry of Exonerations) more
- When deciding whether to pursue the case, the prosecutor must consider the possibility that seemingly convincing evidence is actually *fake*.
- Affects the entire investigation

Introduction

- Study a novel dynamic model of information acquisition, in which information can be faked strategically.
- Characterize equilibria, study positive and normative distortions in information acquisition due to fabrication
- Study changes to search process that mitigate distortions
- Applications: prosecutorial discretion, venture capital investment, project development, product regulation

Model

- Principal makes a single choice b/w safe, risky action
- Safe action, known payoff $\theta \in (0, 1)$ (principal)
- Risky action, uncertain payoff $\omega \in \{0, 1\}$
- Prior belief $\mu \equiv \Pr(\omega = 1)$
- Principal prefers safe under prior, $\mu < \theta$
- Principal can choose S/R at any time $t \geq 0$
- Game ends when she makes this choice
- May delay in order to acquire more information about ω
- Common discount rate ρ

Model

- Agent has no private info about ω
- Same payoff from risky action, $\omega \in \{0, 1\}$
- Smaller payoff from safe action $\beta \in (0, \mu)$
  - No disagreement ex post, $\beta > 0$
  - Under prior, agent prefers risky, principal prefers safe
  - Also, β not too small (more in a few)
- Agent has privately known type.
- With prob. $\sigma \in (0, 1)$, agent is *manipulative* else *normal*
- Significance of agent type clear soon.

Model

- Information about ω comes from a public news process.
- Single arrival of news at some time.
- Arrival may be *real news* or *fake news*.

Model

Real news reports the payoff of the risky action

- Real news is type-1 iff $\omega = 1$
- Real news is type-0 iff $\omega = 0$
- Arrival time of real news uncertain.
- Arrival time drawn from continuous CDF $G(\cdot)$, density $g(\cdot)$, decreasing hazard rate $H_R(\cdot)$.
- Arrival time of real news independent of ω (and A's type)
- If arrival time reached, real news instantly produced

Model

Fake news produced strategically to influence principal

- Manipulative agent can fake a type-1 arrival at any time.
- Fake arrival **looks real**...
- Principal cannot observe or verify if type-1 news is real, she can only infer this from manipulative agent's strategy.
- Normal agent simply waits for game to end.
- Single arrival of news—decision to fake "irreversible."
- If officer decides to plant evidence against suspect, stops looking for alibi

Model

Three helpful observations

1. Only one news arrival; instant decision after arrival.
2. Type-0 not faked. Principal selects safe.
3. News arrival time independent of ω.
   - Real news arrival time independent of ω
   - Fake news arrival time independent of ω (the agent is uninformed)
   - Nice feature: arrival time conveys no info about risky payoff, non-arrival has no effect on either player's belief about it

Model

Autarky Benchmarks

- Autarky: all news is real, relevant player has authority over action/time.
- Each player follows the recommended action when news arrives
- Each player chooses how long to wait before selecting the "default action," safe for principal, risky for agent
- Each player's payoff continuous, differentiable, single-peaked in waiting time

Model

Autarky Benchmarks.

- Optimal to wait until hazard rate reaches a threshold,

$$H_R(\tau_P) = \frac{\rho\theta}{\mu(1-\theta)} \qquad H_R(\tau_A) = \frac{\rho\mu}{\beta(1-\mu)}$$

- Numerator is marginal cost of delaying default action
- Denominator is expected net benefit if default action overturned by news
- Both players want to search, default can be proved wrong
- Focus: $0 < \tau_P < \tau_A$, agent duration longer than principal
- $\tau_P < \tau_A \iff \beta$ not too small, i.e., $\beta \in (\underline{\beta}, \mu)$
- Ensures eq. shaped by disagreement over default action, streamlines analysis

Model

Agent Strategy.

- Agent pure strategy is "faking time," $t \in \mathbb{R}_+$.
- If faking time reached with no decision, agent fakes an arrival at t.
- Agent can mix over faking times, CDF $F_A(\cdot)$.

Model

Principal Strategy.

- Function $a(\cdot)$, probability of risky action if type-1 news arrives at time t.
- Pure strategy, "exit time," $t \in \mathbb{R}_+$
- If exit time reached without news, stop search and select safe action
- Principal can mix over exit times, CDF $F_P(\cdot)$

Model

Payoffs.

- $u_A(t)$ is expected payoff of faking time t, given $F_P(\cdot)$ and $a(\cdot)$
- $u_P(t)$ is expected payoff of exit time t, given $F_A(\cdot)$ and $a(\cdot)$
- Big integrals agent principal
- Key point: distortions from autarky
  - Principal: news might be fake.
  - Agent: principal might exit without news, safe following type-1 news

Equilibrium

Equilibrium Conditions (BNE)

- *Agent Faking*: $f_A(t) > 0 \Rightarrow t \in \text{argmax}_x u_A(x)$
- *Principal Exit*: $f_P(t) > 0 \Rightarrow u_P(t) \in \text{argmax}_x u_P(x)$
- *Principal Action*: $a(t)$ optimal given posterior belief $\mu_1(t)$ (all times).
- *Consistency*: $\mu_1(t)$ from Bayes' rule and agent strategy

Two varieties of equilibrium

1. Beneficial search, $u_P^* \geq \theta$ (focal)
2. Non-beneficial search, $u_P^* = \theta$

Equilibrium Structure

- Suppose principal "naive search," exit at τ_P and $a(\cdot)$=1
- At $t < \tau_P$, no distortions for agent.
- Agent payoff $u_A(t)$, same as first best, increasing for $t < \tau_P < \tau_A$
- Agent doesn't want to fake at $t < \tau_P$
- Principal doesn't want to exit at $t < \tau_P$
- Virtuous Cycle: No faking $\iff$ No early exit
- ... breaks down at τ_P

Equilibrium Structure

- In naive search principal selects *safe* at τ_P without news
- Without news agent prefers *risky*
- Since $a(\cdot) = 1$, agent fakes type-1 news at τ_P to preempt safe
- If such faking expected by the principal, would ignore it, selecting safe. Agent preempts earlier.
- Agent preemption unravels search backwards from τ_P...

Equilibrium Structure

- Two small adjacent time periods, "Early" and "Late," both before τ_P
- "Within period" agent decides whether to fake first.
- If no news arrives principal decides whether to exit, pick safe.
- If news, principal decides whether to follow it.
- Consider best responses loosely

		P	
		Exit Early	Exit Late
A	Fake Early	.. , ..	.. , ..
	Fake Late	.. , ..	.. , ..

Equilibrium Structure

- Agent wants to preempt the principal's choice of safe
- Wants to fake "just before" principal exits/picks safe

		P	
		Exit Early	Exit Late
A	Fake Early	* , ..	.. , ..
	Fake Late	.. , ..	* , ..

Equilibrium Structure

- Principal considers value of future news
- If agent fakes late, then type-1 news in late period most likely fake. Whether arrives or not, picks safe. Not worth waiting for. Fake late $\Rightarrow$ exit early.
- If agent fakes early and no arrival, then agent normal. Late news is real, more valuable, worth waiting for. Fake early $\Rightarrow$ Exit late
- Incentives resemble matching pennies...

		P	
		Exit Early	Exit Late
A	Fake Early	* , ..	.. , *
	Fake Late	.. ,*	* , ..

Equilibrium Structure

- Search unravels *stochastically* from τ_P
- Atoms of stopping and faking at τ_P under naive search "spread out" into interval $[\tau_M, \tau_P]$, with smooth mixing
- Beneficial search cannot unravel to 0, else $u_P^* = \theta$
- Unraveling cannot leave atom fakes at τ_P, initial trigger for unraveling
- Can be atom of stops at τ_P. Agent preempts atom with probability 1

Equilibrium Structure

Proposition 1. If an equilibrium with beneficial search exists, then it has following structure. There exists $\tau_M > 0$ such that

(i) the agent's faking time is drawn from a mixed strategy with no mass points or gaps, supported on $[\tau_M, \tau_P]$.

(ii) the principal's exit time is drawn from a mixed strategy with no gaps, supported on interval $[\tau_M, \tau_P]$; only mass point on τ_P.

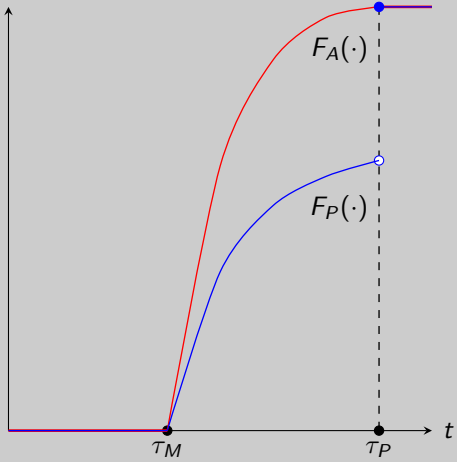
(iii) the principal selects risky following type-1 news, $a(t) = 1$ for all $t \geq 0$.

Equilibrium Structure

Positive Implications

- *Hard deadline*, τ_P. If reached principal exits, selects safe.
- *Soft deadline*, τ_M. If reached, agent randomly fakes, and principal randomly exits in absence of news.
- Principal "disengages" from search at soft deadline; might exit any moment
- Agent becomes "anxious" at soft deadline that principal might exit and pick safe. Fakes news in order to preempt it.
- Even though principal disengaged, acts on type-1 news if arrives
- Future news informative enough to offset waiting cost. If not, stop now!
- To offset waiting cost, future news must be informative enough to follow.

Equilibrium Structure



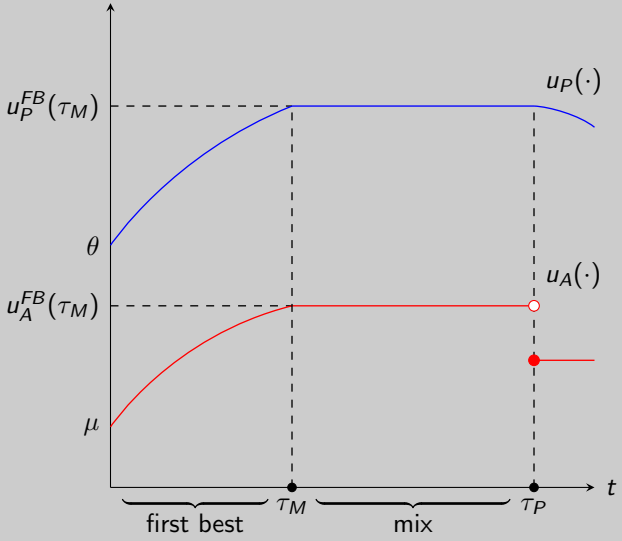
Equilibrium Structure

Equilibrium Structure

Normative Implication

- Before τ_M, no faking/exit—like first best search
- In equilibrium players indifferent over all times in $[\tau_M, \tau_P]$.
- Equilibrium payoff as if each player exits/fakes at τ_M.
- Principal's payoff as if waits for real news until τ_M, then picks safe
- Agent's payoff as if waits for real news until τ_M, then picks risky
- As if each player does an autarkic search but cuts too early.

Equilibrium Structure



Equilibrium payoffs, Beneficial Search

Equilibrium Characterization

Beneficial Search (Proposition 2)

- Equilibrium with beneficial search exists if $\sigma < \bar{\sigma}(\theta, \mu)$.
- At most one equilibrium with beneficial search exists
- Closed form characterization see it

Beneficial Search Equilibrium

Notable Features

- Always atom on τ_P for principal.
- Atom and $a(\cdot) = 1$, positive probability principal acts "naively"
- Informativeness of type-1 news non-monotone in time. Big drop at τ_M, recovers gradually, restored at τ_P
- Increase in σ intensifies unraveling, reduces soft deadline, hurts both players
- If too big ($\sigma > \bar{\sigma}$), unravels to 0. No beneficial search.

Remedies

Commitment to Naive Search

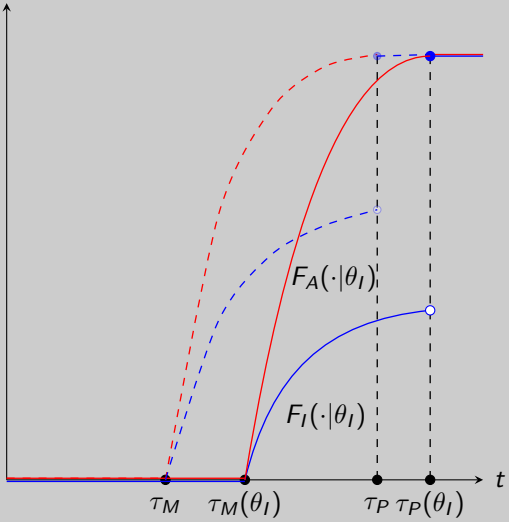
- Suppose principal *commits* to act naively: select safe at τ_P if no news, act on all type-1 news that arrives, $a(\cdot) = 1$.
- Manipulative agent waits to fake until τ_P
- No preemption incentive for agent. No unraveling (Yay!)
- But, if agent is manipulative, fakes type-1 at τ_P. Principal picks risky instead of safe (Boo!)
- Result: compared to beneficial search equilibrium, commitment to naive search generates improvement
- "manipulative agent unlikely, let's pretend doesn't exist"
- P benefits from "plausible deniability" of faking, without accountability harmed by scrutiny

Remedies

Delegation to Intermediary

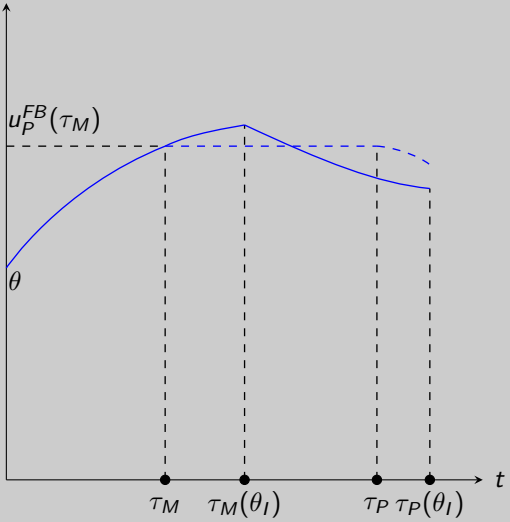
- More mild form of delegation: principal introduces an intermediary who has full authority over search/action
- Intermediary payoff from safe action θ_I (local)
- Game between intermediary and agent.
- Focus on (more interesting) case of beneficial search
- Smaller θ_I aligns incentives better, equilibrium "shifts up"
- Higher τ_M (Yay! more first best search)
- Intermediary has lower value of stopping than principal. If intermediary indifferent, principal wants to stop. Principal payoff decreasing over intermediary support (Boo!)

Delegating Down



Eq. Strategies, Delegation

Delegating Down



Delegation Payoff

Remedies

Delegation to Intermediary

- Result: compared to keeping authority herself, principal can benefit by delegating to an intermediary with a smaller θ_I
- Delegating charging decision to grand jury can be helpful, if less-concerned about prosecutor's opportunity costs
- Delegating to an expert who is a bit more concerned with "long term viability" than opportunity cost benefits VC.
- (P tempted to stop search, not overrule action)

Thanks for your attention!

Agent Payoff

- Consider $u_A(t)$ agent's payoff from faking time t.
- Four ways the game can end

Time	Event	Prob/density	Payoff	Discount
$s < t$	Real Type-0	$w_0^A(s) \equiv (1-\mu)g(s)(1-F_P(s))$	0	$\exp(-\rho s)$
$s < t$	Real Type-1	$w_1^A(s) \equiv \mu g(s)(1-F_P(s))$	$\beta(1-a(s))+a(s)$	$\exp(-\rho s)$
$s < t$	P Stops	$w_S^A(s) \equiv f_P(s)(1-G(s))$	β	$\exp(-\rho s)$
t	A fakes	$W_\phi^A(t) \equiv (1-G(t))(1-F_P(t))$	$\beta(1-a(t))+\mu a(t)$	$\exp(-\rho t)$

$$u_A(t) = \int_0^t \exp(-\rho s)\{w_0^A(s)\beta + w_1^A(s)((1-a(s))\beta + a(s)) + w_S^A(s)\beta\}ds$$
$$+ \exp(-\rho t)W_\phi^A(t)((1-a(t))\beta + \mu a(t)).$$

- $u_A(t)$ is expectation, based on above.
- Note that agent payoff different when real type-1 vs. fake. back

Principal Payoff

- Consider $u_P(t)$ agent's payoff from stopping time t.

Time	Event	Prob/density	Payoff	Discount
$s < t$	Real Type-0	$w_0^P(s) \equiv (1-\mu)g(s)(1-\sigma F_A(s))$	0	$\exp(-\rho s)$
$s < t$	Real Type-1	$w_{1R}^P(s) \equiv \mu g(s)(1-\sigma F_A(s))$	$\theta(1-a(s)) + a(s)\mu_1(s)$	$\exp(-\rho s)$
$s < t$	Fake Type-1	$w_{1F}^P(s) \equiv \sigma f_P(s)(1-G(s))$	$\theta(1-a(s)) + a(s)\mu_1(s)$	$\exp(-\rho s)$
t	P Stops	$W_\phi^P(t) \equiv (1-G(t))(1-\sigma F_A(t))$	θ	$\exp(-\rho t)$

- Principal cannot observe whether type-1 is real or fake. Both type-1's in a single information set.
- $w_1^P(s) = w_{1R}^P(s) + w_{1F}^P(s)$ and $\mu_1(s) \equiv \Pr(\omega = 1 |$ type-1 at $s) = w_{1R}^P(s)/w_1^P(s)$.

$$u_P(t) = \int_0^t \exp(-\rho s)\{w_0^P(s)\theta + w_1^P(s)((1-a(s))\theta + a(s)\mu_1(s))\}ds + \exp(-\rho t)W_\phi^P(t))\theta.$$

- $u_P(t)$ is expectation, based on above. back

Equilibrium Characterization

$$F_A(t) = \frac{1}{\sigma}\Big(1 - \exp\{-\int_{\tau_M}^{t} \frac{\mu(1-\theta)H_R(s) - \rho\theta}{\theta - \mu} ds\}\Big)$$

$$t \in [\tau_M, \tau_P) \Rightarrow F_P(t) = 1 - \exp\{-\int_{\tau_M}^{t} \frac{\beta(1-\mu)H_R(s) - \rho\mu}{\mu - \beta} ds\},$$

$$F_P(\tau_P) = 1.$$

τ_M is unique solution to

$$1 - \exp\{-\int_{\tau_M}^{\tau_P} \frac{\mu(1-\theta)H_R(s) - \rho\theta}{\theta - \mu} ds\} = \sigma.$$

Note $\tau_M > 0$ if $\sigma < \bar{\sigma}$,

$$\bar{\sigma} \equiv 1 - \exp\{-\int_{0}^{\tau_P} \frac{\mu(1-\theta)H_R(s) - \rho\theta}{\theta - \mu} ds\}.$$

back

Police Misconduct

In a widening scandal that has rocked the New York State Police, a lieutenant who supervised criminal investigations in seven upstate counties admitted yesterday that he had faked fingerprint evidence in three cases...

Taken together, yesterday's events painted a picture of **almost routine fabrication** *of evidence in criminal cases*

—*New York Times*, July 30 1993

Police Misconduct

In any reasonable mind, a serious question of how that stain and single hair came to be found in the car is raised... the possibility of [the evidence being planted] is very real and raises doubts about the credibility of the evidence and the police

—*West Virginia Supreme Court*, July 1992 back