Privacy Protection: When Does Hiding in Plain Sight Work?

Tatiana Mayskaya[1] Arina Nikandrova[2]

[1]Higher School of Economics

[2]City, University of London

ITAM
22 November 2019

Privacy Debate

Protection against access of personal information:

- general protection (all information) — government responsibility
- selective protection (only sensitive information) — individual responsibility

Goal: protection of sensitive information

✗ General protection has high indirect cost since it limits access to big data

> *Your individual data is actually not that valuable. While the entire data market might be worth $3trn... it's access to huge aggregate data that is valuable.*
>
> Privacy International

✓ Could providing tools for selective protection be a solution?

Hiding in Plain Sight

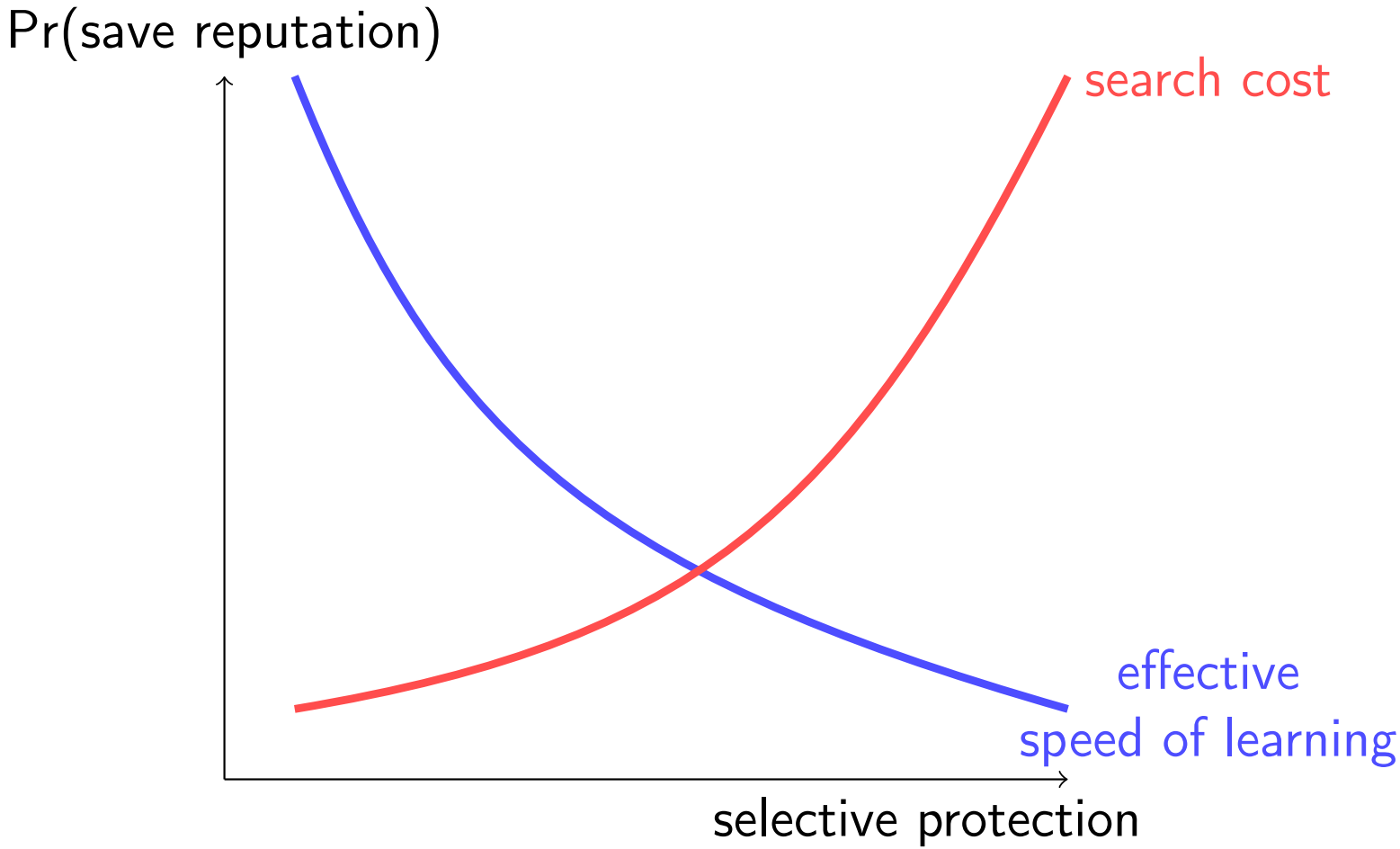
Take-away

When general protection is imperfect and perfect selective protection is infeasible, low selective protection ("hiding in plain sight") becomes the optimal strategy for an individual even in the absence of protection cost.

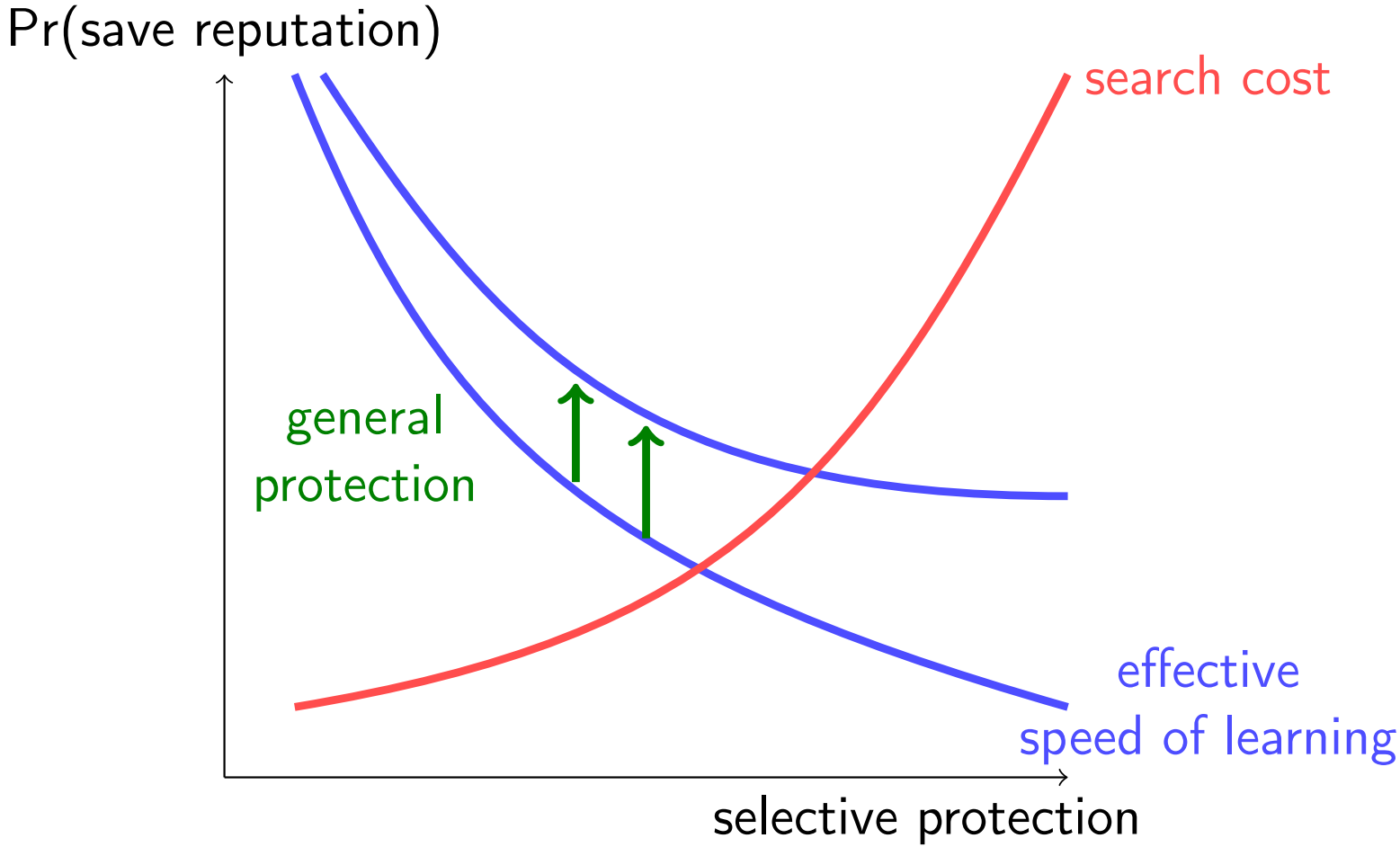
Policy implication

Tools that facilitate selective protection might not be used in practice in the absence of good general protection.

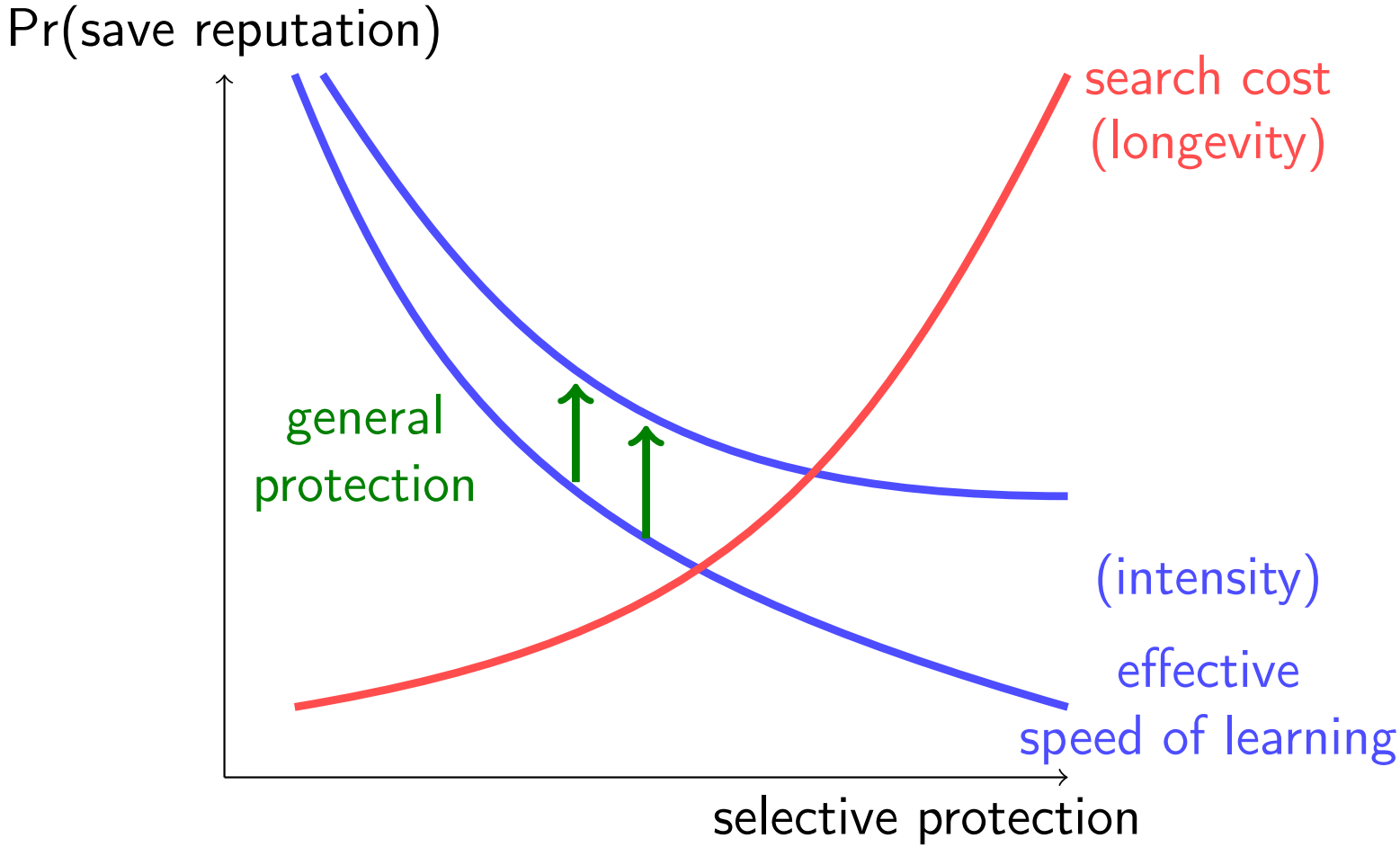
Trade-Off


Pr(save reputation)
search cost
effective
speed of learning
selective protection

Trade-Off


Pr(save reputation)
search cost
general
protection
effective
speed of learning
selective protection

Trade-Off



Trade-off: *intensity* vs *longevity*

- strong protection ⇒ hard to find ⇒ low longevity of search
- weak protection ⇒ quickly become pessimistic about finding anything ⇒ high intensity of learning

Outline

Model

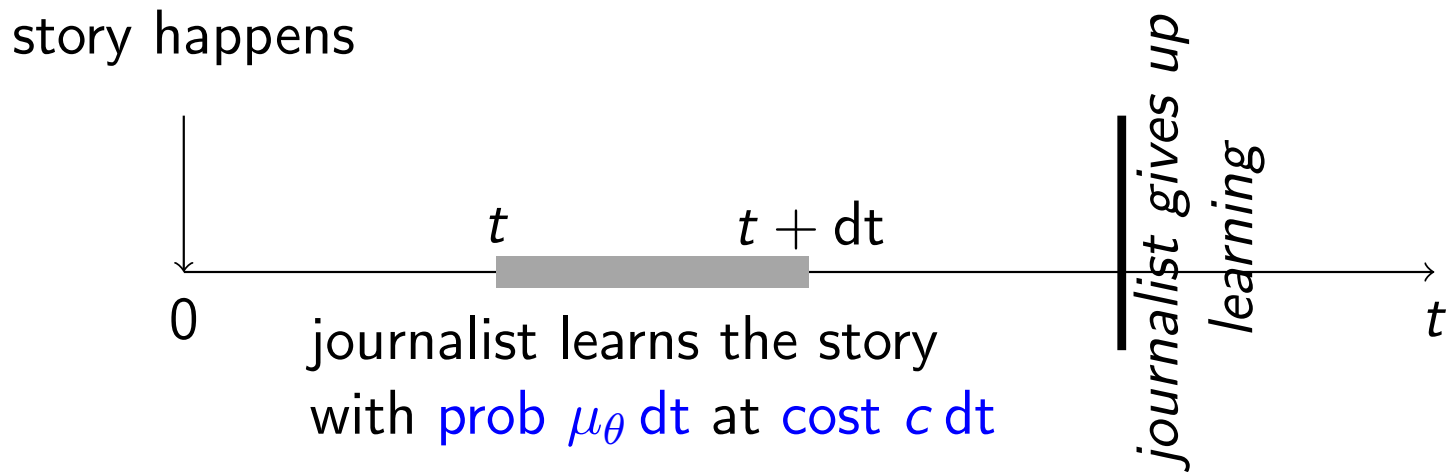
- The game is between a celebrity and a journalist.
- The celebrity publicly commits to the level of selective protection parametrized by $\mu_1 > 0$.
  - useful connections, advocates, loopholes in protocols, options to remove parts of own digital fingerprint, installed VPN service, etc $\Rightarrow$ tools that the celebrity would use if and only if a story that compromises her happens

Model

- The game is between a celebrity and a journalist.
- The celebrity publicly commits to the level of selective protection parametrized by $\mu_1 > 0$.
- A story that draws the journalist's attention happens. It could be either compromising ($\theta = 1$) or not interesting ($\theta = 0$). The probability it is compromising is $p \in (0, 1)$.
  - exogenous $\Rightarrow$ independent of ex post protection; no ex ante signaling through protection level
  - interpretations of p: publicly observable ability of a journalist to detect an interesting story; type of the celebrity reflecting her propensity to get involved in a scandal

Model

- The game is between a celebrity and a journalist.
- The celebrity publicly commits to the level of selective protection parametrized by $\mu_1 > 0$.
- A story that draws the journalist's attention happens. It could be either compromising ($\theta = 1$) or not interesting ($\theta = 0$). The probability it is compromising is $p \in (0, 1)$.



- $\mu_0 > 0$ characterizes general protection
- If the journalist knows the story, he can report it. Once he reports the story he gets $\beta > 0$ if $\theta = 1$ and some negative payoff otherwise.
- The celebrity minimizes the probability of a report

Beliefs

As long as the journalist searches and does not find the story, his belief p_t about $\theta = 1$

(D) drifts down when $\mu_1 > \mu_0$ (selective protection is less than general protection)

(C) stay constant when $\mu_1 = \mu_0$

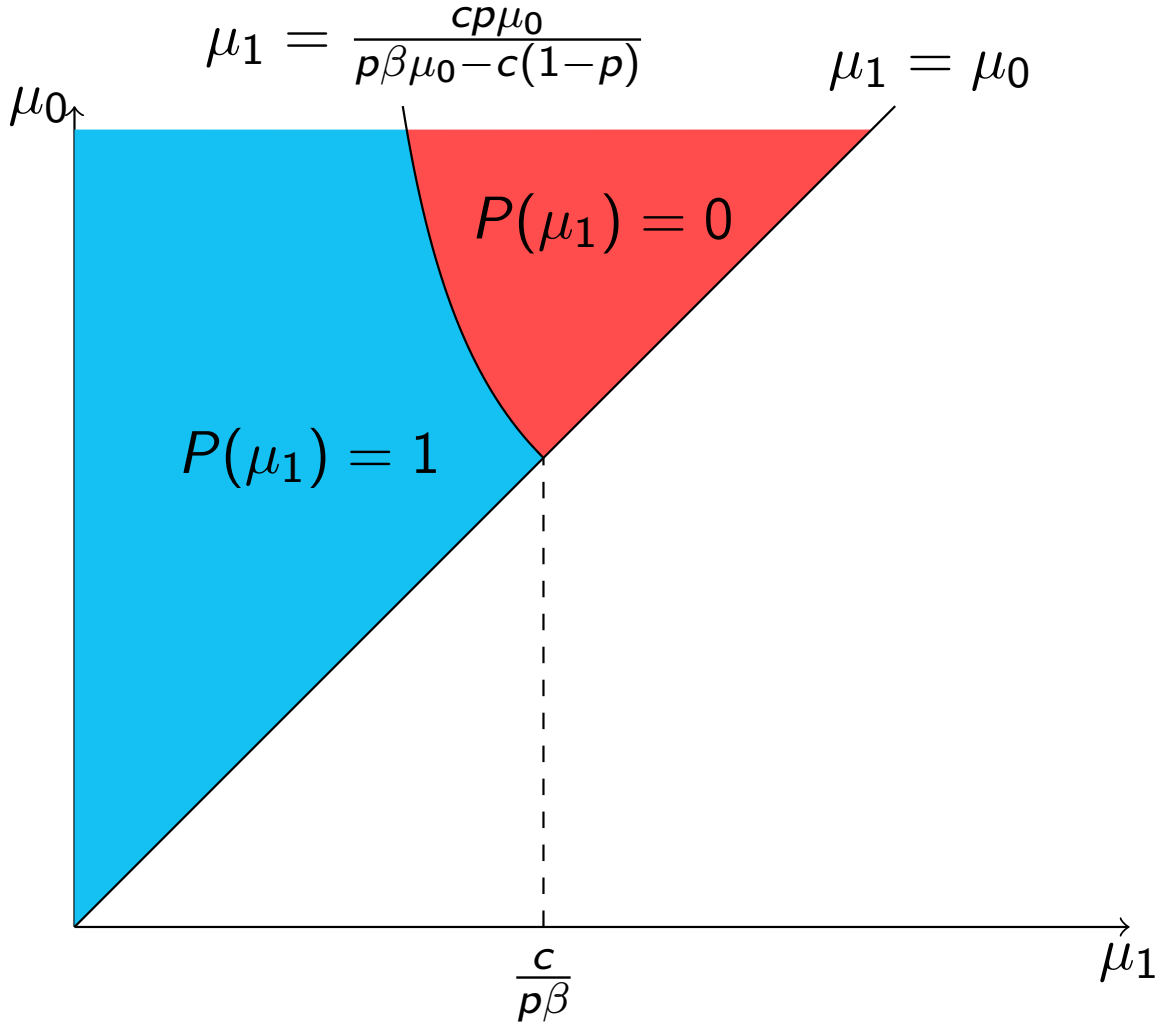
(U) drifts up when $\mu_1 < \mu_0$

$$\left(\ln \frac{p_t}{1 - p_t} \right)'_t = \mu_0 - \mu_1$$

(C)+(U): Expected benefit from learning until the story is found:

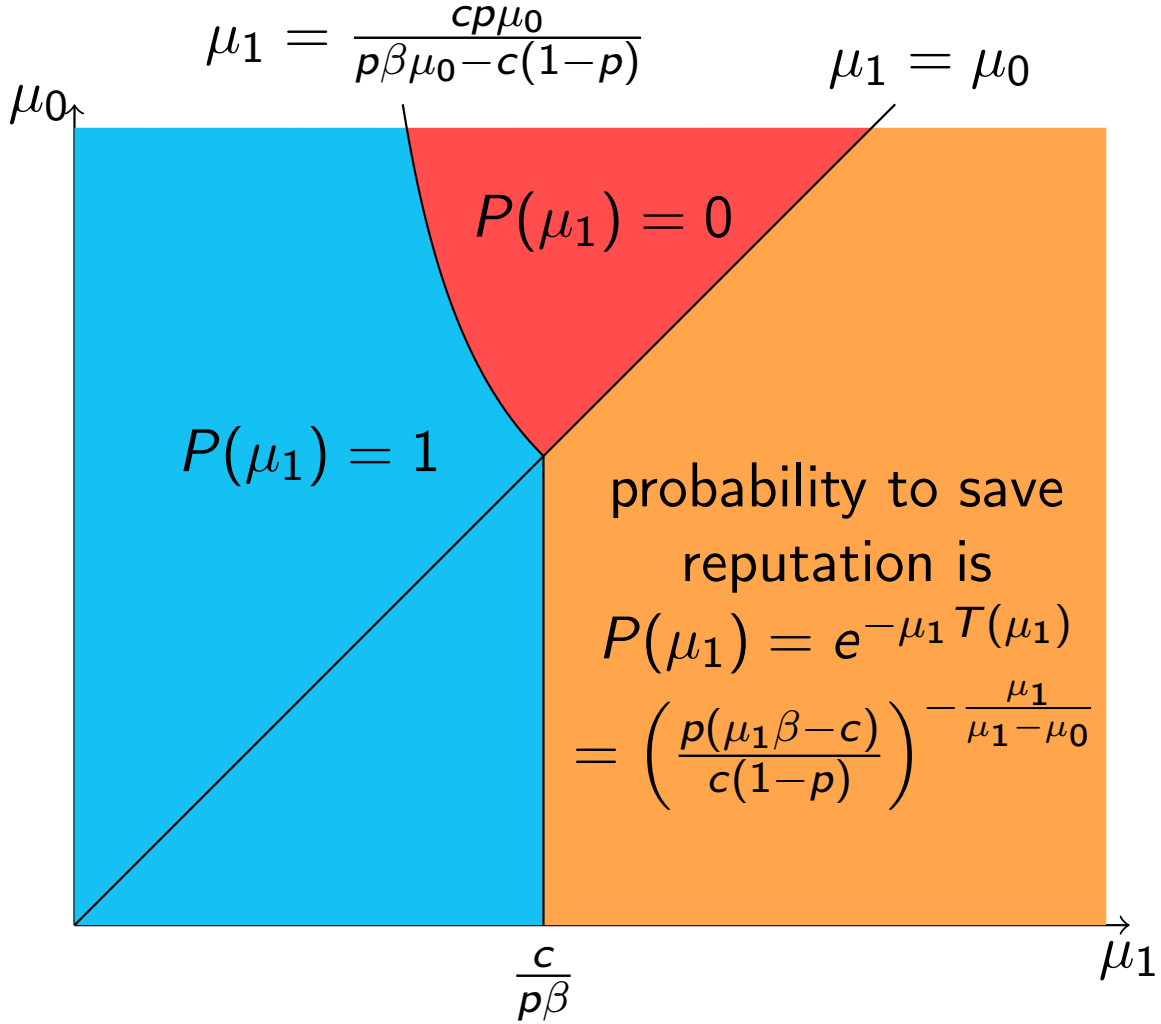
$$V(\mu_1) = p\beta - \frac{cp}{\mu_1} - \frac{c(1-p)}{\mu_0}$$

Full protection $\Leftrightarrow V(\mu_1) \leq 0 \Leftrightarrow \mu_1 \left(p\beta\mu_0 - c(1-p)\right) \leq cp\mu_0$.

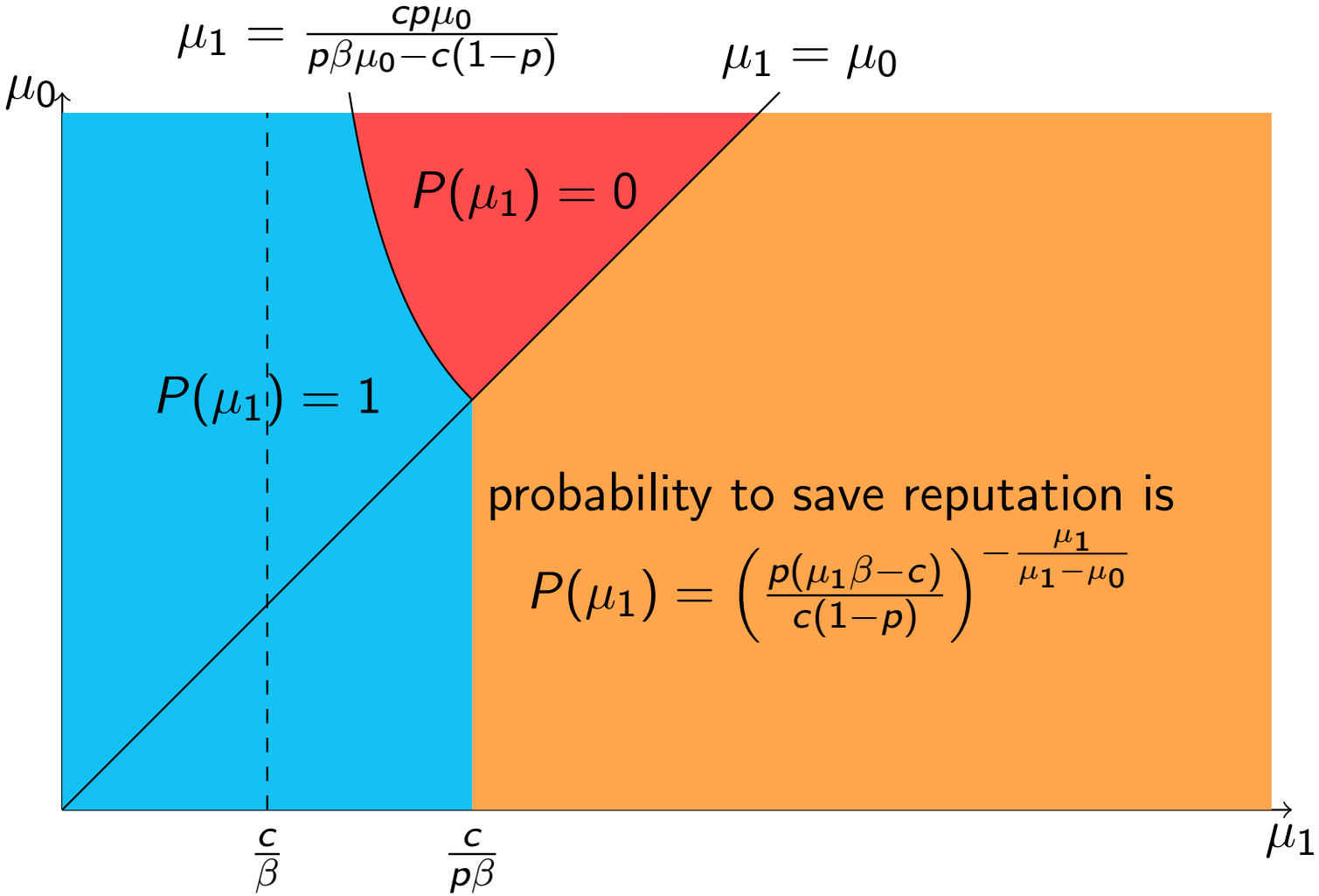


(D): The optimal stopping time in the absence of finding:

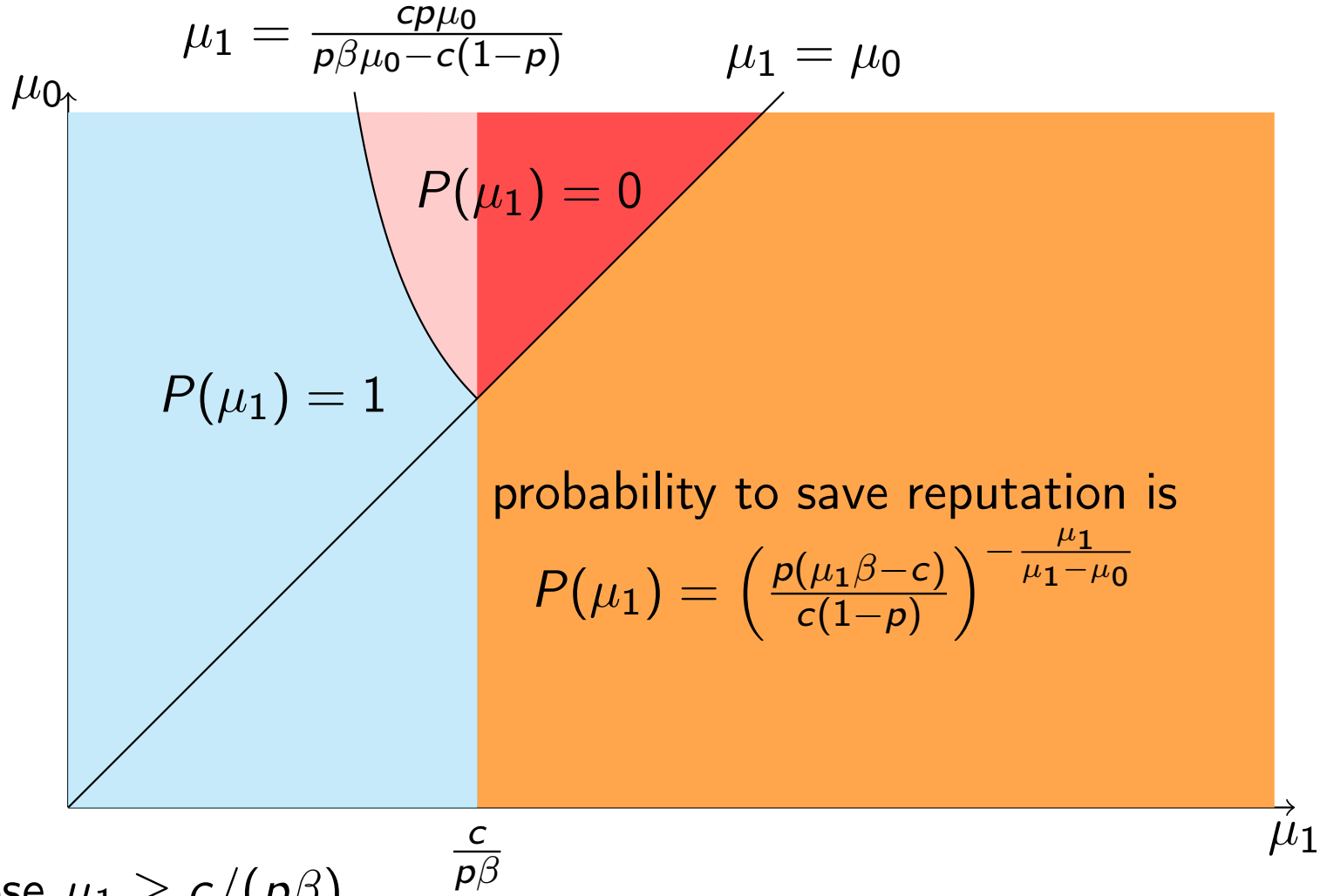
$$T(\mu_1) = \begin{cases} \frac{1}{\mu_1 - \mu_0} \ln\left(\frac{p(\mu_1\beta - c)}{c(1-p)}\right), & \mu_1 > \frac{c}{p\beta} \\ 0, & \mu_1 \leq \frac{c}{p\beta} \end{cases}$$



Optimal Selective Protection

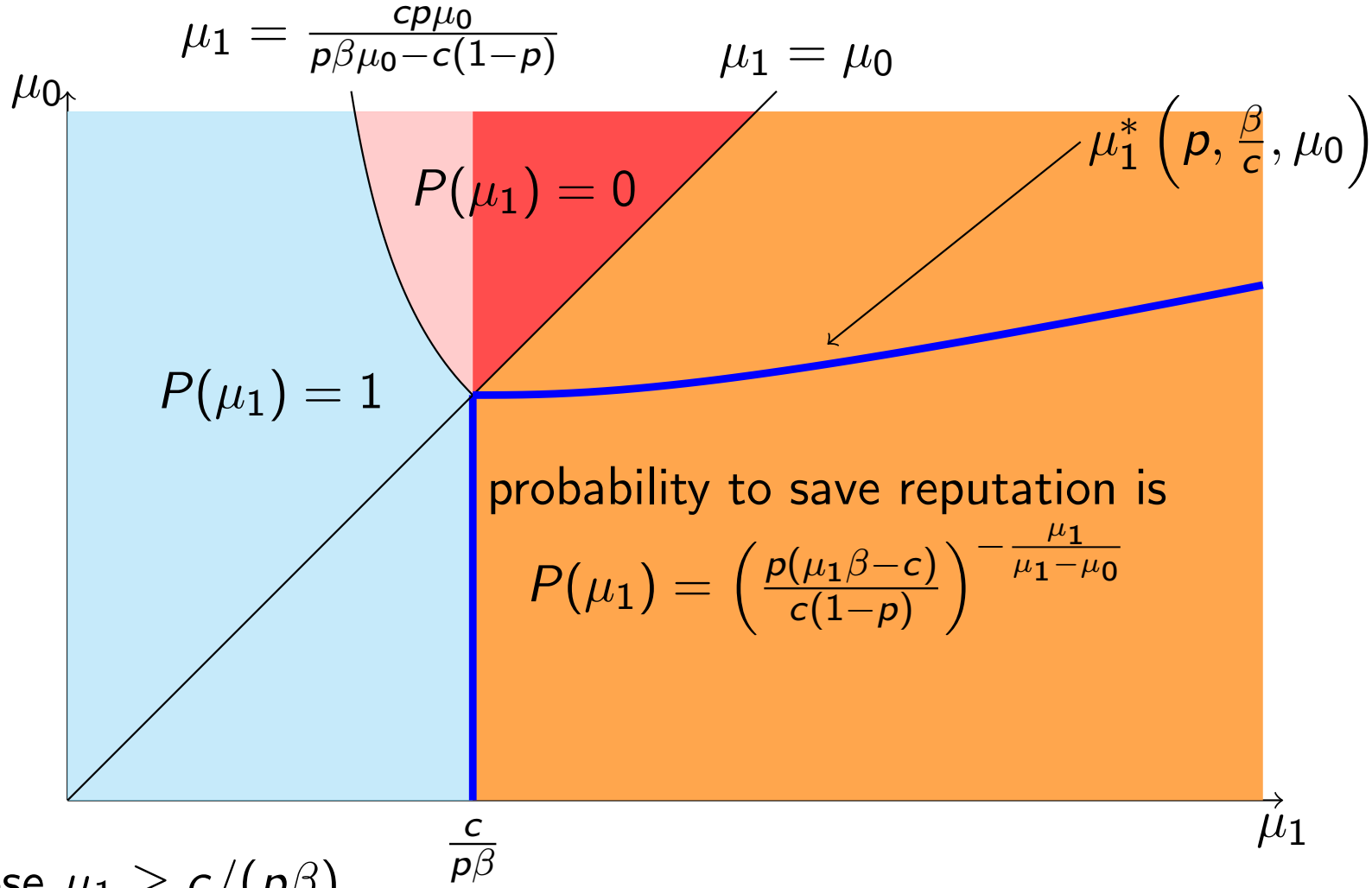


Optimal Selective Protection



suppose $\mu_1 \geq c/(p\beta)$

Optimal Selective Protection



suppose $\mu_1 \geq c/(p\beta)$

- if $\mu_0 < c/(p\beta)$, then $\mu_1^* = c/(p\beta)$
- if $\mu_0 > c/(p\beta)$, then $\mu_1^*(p, \beta/c, \mu_0) > \mu_0$

Trade-Off

$$P(\mu_1) = \left(\frac{p(\mu_1 \beta - c)}{c(1-p)} \right)^{-\frac{\mu_1}{\mu_1 - \mu_0}}$$

- search cost (longevity of search):

$$\uparrow c \sim \downarrow \mu_1 \quad \Rightarrow \quad \uparrow \underline{p} = \min \left\{ \frac{c}{\mu_1 \beta}, p \right\} \quad \Rightarrow \quad \uparrow P(\mu_1)$$

- effective speed of learning (intensity of learning):

$$\dot{q}_t = \left(\ln \frac{1 - p_t}{p_t} \right)'_t = \mu_1 - \mu_0$$

$$\downarrow \mu_1 \quad \Rightarrow \quad \downarrow \frac{q_{t+dt} - q_t}{\mu_1 \, \mathrm{dt}} = \frac{\dot{q}_t}{\mu_1} = 1 - \frac{\mu_0}{\mu_1} \quad \Rightarrow \quad \downarrow P(\mu_1)$$

Trade-Off

$$P(\mu_1) = \left(\frac{p(\mu_1\beta - c)}{c(1-p)} \right)^{-\frac{\mu_1}{\mu_1 - \mu_0}}$$

- search cost (longevity of search):

$$\uparrow c \sim \downarrow \mu_1 \quad \Rightarrow \quad \uparrow \underline{p} = \min\left\{ \frac{c}{\mu_1\beta}, p \right\} \quad \Rightarrow \quad \uparrow P(\mu_1)$$

- effective speed of learning (intensity of learning):

$$\dot{q}_t = \left(\ln \frac{1-p_t}{p_t} \right)'_t = \mu_1 - \mu_0$$

$$\downarrow \mu_1 \quad \Rightarrow \quad \downarrow \frac{q_{t+dt} - q_t}{\mu_1 \, dt} = \frac{\dot{q}_t}{\mu_1} = 1 - \frac{\mu_0}{\mu_1} \quad \Rightarrow \quad \downarrow P(\mu_1)$$

Intuition:

- $\mu_0 = \mu_1 \Rightarrow \dot{q}_t/\mu_1 = 0$
- $\dot{q}_t/\mu_1 \downarrow$ in μ_0

$\Rightarrow \dot{q}_t/\mu_1 < 0$ for $\mu_0 > \mu_1$

$\Rightarrow \dot{q}_t/\mu_1 \uparrow$ in μ_1 in some neighborhood of μ_0

Comparative Statics

Prior Belief

$\mu_1^*\left(p, \frac{\beta}{c}, \mu_0\right)$ is increasing in p

- $\uparrow$ $p \Rightarrow \uparrow$ learning duration $\Rightarrow$ *intensity of learning* becomes more important $\Rightarrow \uparrow \mu_1$ to increase the *intensity of learning*

Comparative Statics

Prior Belief

$\mu_1^*\left(p, \frac{\beta}{c}, \mu_0\right)$ is increasing in p

- $\uparrow$ $p \Rightarrow \uparrow$ learning duration $\Rightarrow$ *intensity of learning* becomes more important $\Rightarrow \uparrow \mu_1$ to increase the *intensity of learning*

Benefit/Cost

$\mu_1^*\left(p, \frac{\beta}{c}, \mu_0\right)$ is increasing in $\frac{\beta}{c}$

- $\uparrow$ $\beta/c \Rightarrow \uparrow$ learning duration $\Rightarrow$ *intensity of learning* becomes more important $\Rightarrow \uparrow \mu_1$ to increase the *intensity of learning*

Comparative Statics

Prior Belief

$\mu_1^*\left(p, \frac{\beta}{c}, \mu_0\right)$ is increasing in p

- $\uparrow\ p \Rightarrow \uparrow$ learning duration $\Rightarrow$ *intensity of learning* becomes more important $\Rightarrow \uparrow \mu_1$ to increase the *intensity of learning*

Benefit/Cost

$\mu_1^*\left(p, \frac{\beta}{c}, \mu_0\right)$ is increasing in $\frac{\beta}{c}$

- $\uparrow\ \beta/c \Rightarrow \uparrow$ learning duration $\Rightarrow$ *intensity of learning* becomes more important $\Rightarrow \uparrow \mu_1$ to increase the *intensity of learning*

General Protection

$\mu_1^*\left(p, \frac{\beta}{c}, \mu_0\right)$ is increasing in μ_0

- $\uparrow\ \mu_0 \Rightarrow \downarrow$ *intensity of learning* $\Rightarrow \uparrow \mu_1$ to increase the *intensity of learning*

Outline

Extension: Many Journalists

- $n \geq 1$ journalists
- fixed μ_1 and μ_0; the celebrity chooses $n \geq 1$ at no cost (control access)
- only the journalist who publishes the compromising story the first gets β (*exclusivity* assumption: information could be used only once)
- independent private learning, reports are public

Learning Pattern

Conditional on no report:

- compromising story

nobody knows the story;
all journalists invest in learning

Learning Pattern

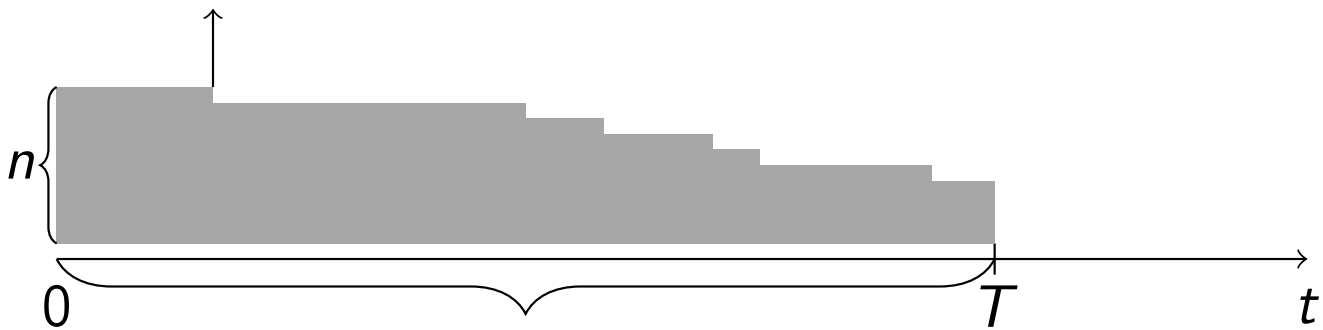
Conditional on no report:

- compromising story

nobody knows the story;
all journalists invest in learning

- non-compromising story

journalist i learns the story

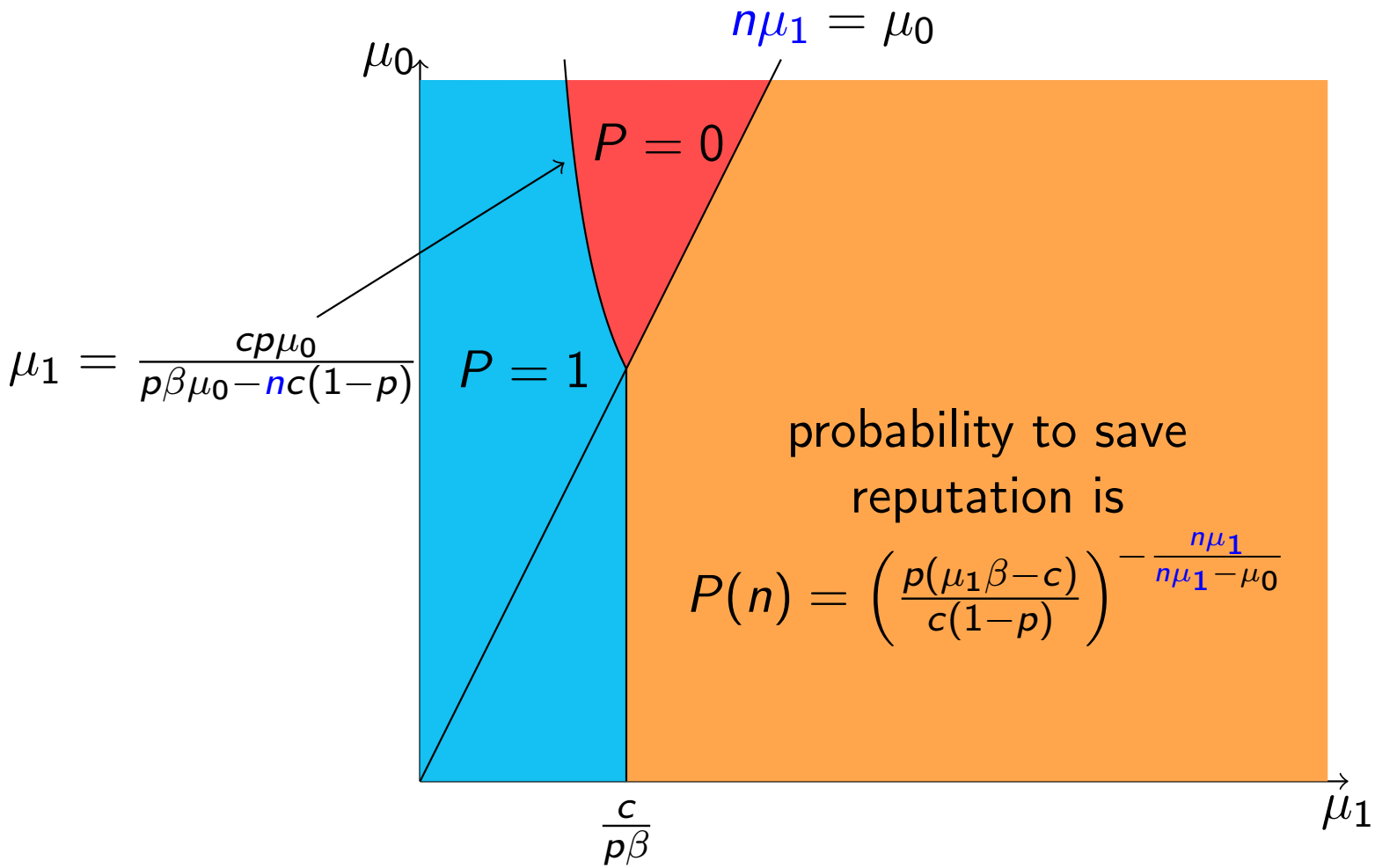


all journalists who do not know the story
invest in learning

Rate of learning for the compromising story is $n\mu_1$ instead of μ_1:

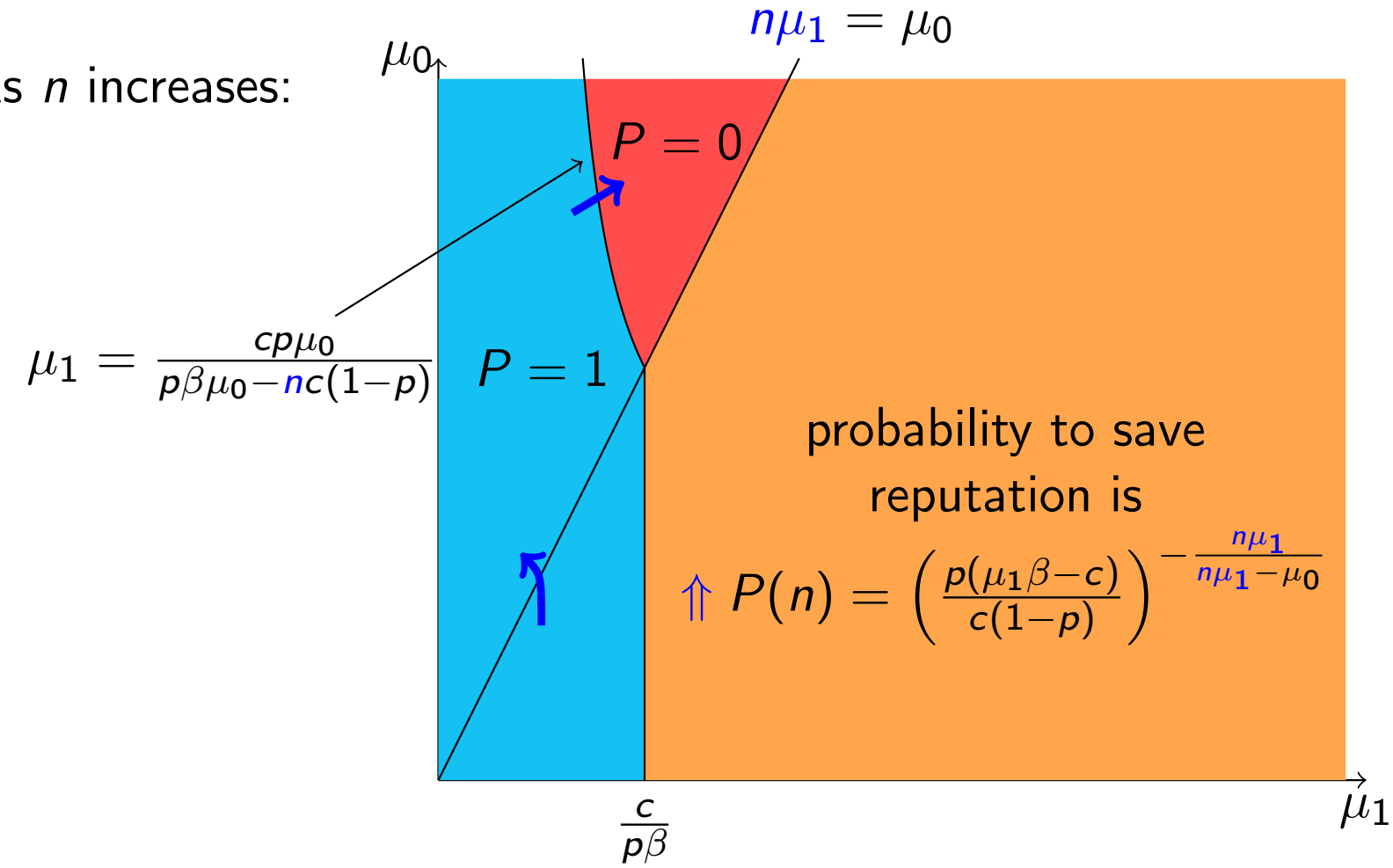
$$\left(\ln \frac{p_t}{1-p_t}\right)'_t = \mu_0 - n\mu_1$$

Yet the stopping threshold is still $\underline{p} = \frac{c}{\mu_1 \beta}$



Optimal Access

As n increases:



Conclusion: $n = +\infty$ (open access) is optimal

Trade-Off

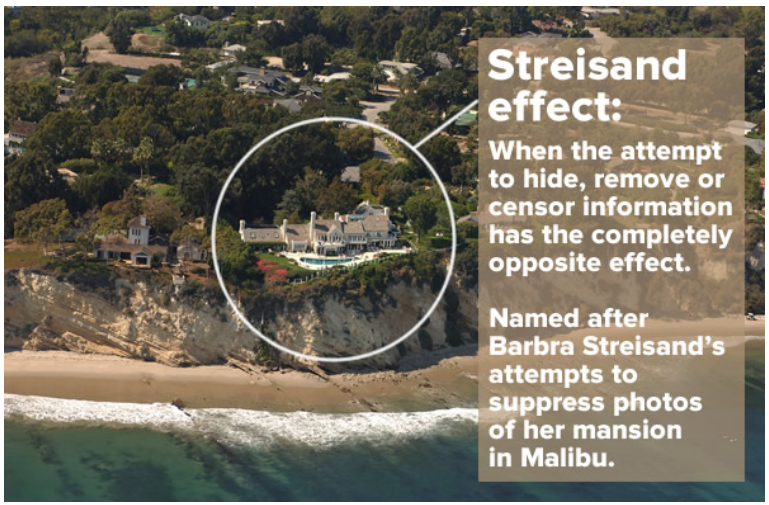
$$P(n) = \left(\frac{p(\mu_1 \beta - c)}{c(1-p)} \right)^{-\frac{n\mu_1}{n\mu_1 - \mu_0}}$$

- search cost (longevity of search): stay the same since stopping threshold $\underline{p} = \min\left\{ \frac{c}{\mu_1 \beta}, p \right\}$ does not depend on n
- effective speed of learning (intensity of learning):

$$\dot{q}_t = \left(\ln \frac{1-p_t}{p_t} \right)'_t = n\mu_1 - \mu_0$$

$$\downarrow n \quad \Rightarrow \quad \downarrow \frac{\dot{q}_t}{n\mu_1} = 1 - \frac{\mu_0}{n\mu_1} \quad \Rightarrow \quad \downarrow P(n)$$

Examples: Streisand Effect



Streisand Effect

Barbra Streisand failed to take into account the indirect cost of increasing selective protection: stronger protection attracts more attention.

Examples: *Twilight* (2008)

Sometimes the best hiding place is the one that's in plain sight.
Stephenie Meyer

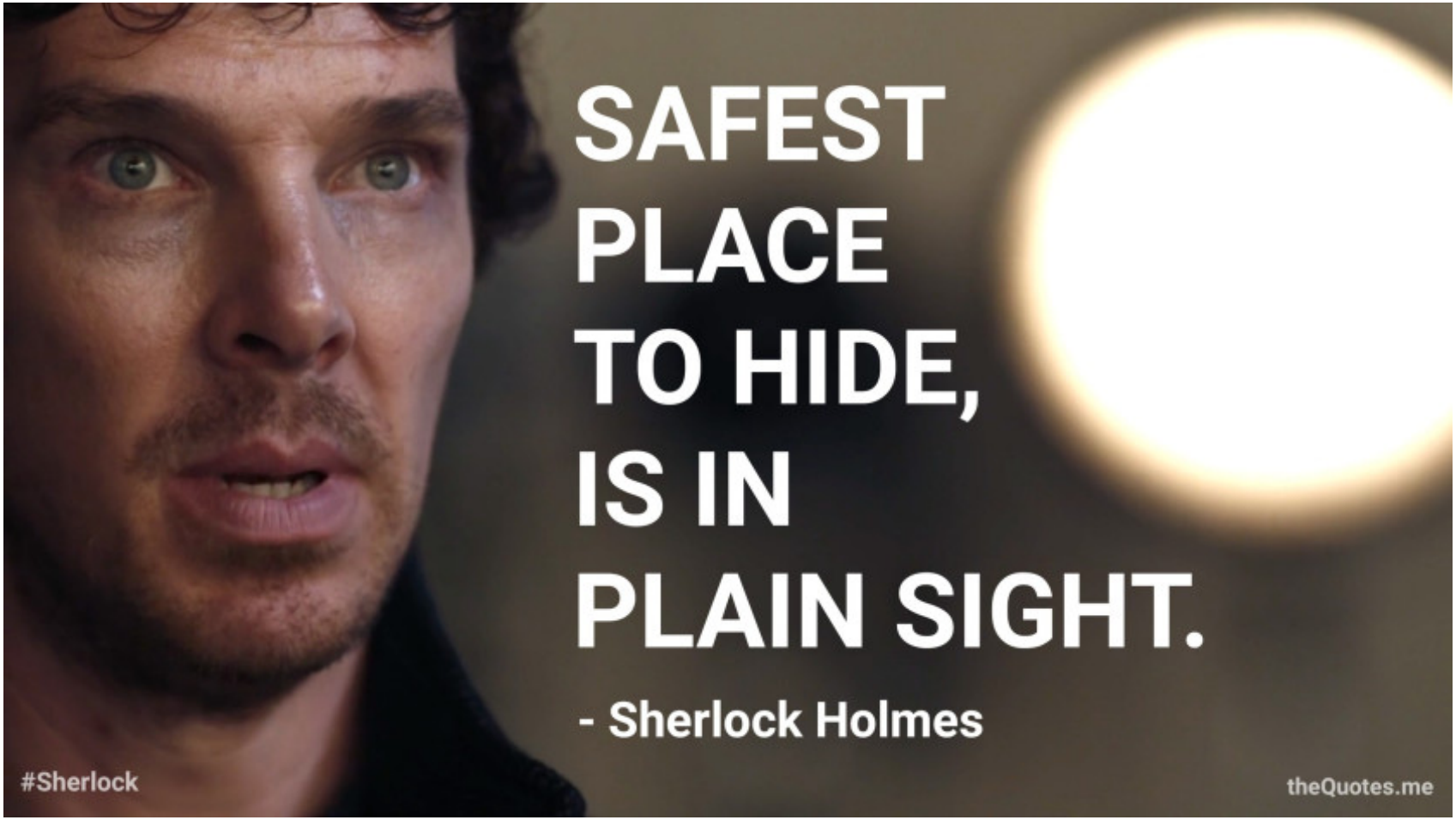
A vampire family was hiding in plain sight (e.g. go to school) all along and nobody figured that out...

Examples: Anti-Corruption Foundation in Russia

- since 2011 ACF accused of corruption more than 50 Russian government officials
- staff includes only 30 people
- have access only to publicly available data

Why does government provide so little protection for their own?

- almost all government officials in Russia are corrupt.
- the most corrupt officials are also members of the elite "club" that has government as their "krysha"
- low "general" protection for government officials outside of the "club" $\Rightarrow$ low protection for their own


SAFEST
PLACE
TO HIDE,
IS IN
PLAIN SIGHT.
- Sherlock Holmes
#Sherlock
theQuotes.me

Literature

- Privacy of consumer data: review by Acquisti et al. (2016)
- Costly maintaining privacy: Conitzer et al. (2012) inverted U-shape relationship between social welfare and privacy costs
- Intrinsic privacy concerns:
  - Gradwohl (2018) decision making in committees
  - Dziuda and Gradwohl (2015) interfirm communication to achieve cooperation
  - Gradwohl and Smorodinsky (2017) signaling games
- Strategic experimentation with Poisson bandits: Keller, Rady and Cripps (2005) and the rich literature that followed
- Wald testing in continuous time with Poisson signals: Peskir and Shiryaev, 2006, ch. VI