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

City, University of London
October 2020

Prince Harry and Meghan Markle: In pursuit of privacy

Prince Harry's lawsuit against tabloids could backfire, commentators claim

Duke of Sussex's legal action against Sun and Daily Mirror over alleged phone hacking takes attack on press up a level

The Guardian, 5 October 2019

Prince Harry and Meghan Markle: In pursuit of privacy

Prince Harry's lawsuit against tabloids could backfire, commentators claim

Duke of Sussex's legal action against Sun and Daily Mirror over alleged phone hacking takes attack on press up a level

The Guardian, 5 October 2019

The Sussexes have reportedly collaborated with a book about their split from the royal family. So much for pursuing a new life of privacy

The Guardian, 5 May 2020

Framework outline

- Two players: a hider (she) and a seeker (he).

Framework outline

- Two players: a hider (she) and a seeker (he).
- Hider may or may not have compromising information to hide.

Framework outline

- Two players: a hider (she) and a seeker (he).
- Hider may or may not have compromising information to hide.
- Seeker aims to find and expose compromising information.

Framework outline

- Two players: a hider (she) and a seeker (he).
- Hider may or may not have compromising information to hide.
- Seeker aims to find and expose compromising information.
- Non-compromising information is protected at some default level:
  - ex: privacy laws, encryption of data in WhatsApp, US tradition of no trespassing.

Framework outline

- Two players: a hider (she) and a seeker (he).
- Hider may or may not have compromising information to hide.
- Seeker aims to find and expose compromising information.
- Non-compromising information is protected at some default level:
  - ex: privacy laws, encryption of data in WhatsApp, US tradition of no trespassing.
- Hider commits to some level of protection for compromising information:
  - ex: Royal family security protocols, one's habits on social media, company's press releases.

Framework outline

- Two players: a hider (she) and a seeker (he).
- Hider may or may not have compromising information to hide.
- Seeker aims to find and expose compromising information.
- Non-compromising information is protected at some default level:
  - ex: privacy laws, encryption of data in WhatsApp, US tradition of no trespassing.
- Hider commits to some level of protection for compromising information:
  - ex: Royal family security protocols, one's habits on social media, company's press releases.
- Technological constraint precludes perfect protection of compromising information.

Framework outline

- Two players: a hider (she) and a seeker (he).
- Hider may or may not have compromising information to hide.
- Seeker aims to find and expose compromising information.
- Non-compromising information is protected at some default level:
  - ex: privacy laws, encryption of data in WhatsApp, US tradition of no trespassing.
- Hider commits to some level of protection for compromising information:
  - ex: Royal family security protocols, one's habits on social media, company's press releases.
- Technological constraint precludes perfect protection of compromising information.
- The realised level of protection is private to the hider.

Results preview

- The hider publicly commits to relatively low protection of compromising information, to persuade the seeker to give up the search quicker.

Results preview

- The hider publicly commits to relatively low protection of compromising information, to persuade the seeker to give up the search quicker.
- The optimal protection level for compromising information is increasing in the default protection, thus making these two types of protection complementary.

Results preview

- The hider publicly commits to relatively low protection of compromising information, to persuade the seeker to give up the search quicker.
- The optimal protection level for compromising information is increasing in the default protection, thus making these two types of protection complementary.
- When the hider can control the number of seekers, the open access policy with infinite number of seekers is optimal.

Why should we care?

Why should we care?

Other examples

- company hiding financial problems
- central bank hiding the depletion of foreign reserves
- politician hiding her misdeeds

Why should we care?

Other examples

- company hiding financial problems
- central bank hiding the depletion of foreign reserves
- politician hiding her misdeeds

Policy relevant question

How should a society design its privacy laws?

Is the default protection a complement or substitute to private efforts to protect sensitive information?

Literature

- Privacy as anonymity to avoid price discrimination: Acquisti, Taylor, and Wagman (2016)
- Intrinsic value of privacy: Gradwohl (2018), Dziuda and Gradwohl (2015), Gradwohl and Smorodinsky (2017)
- Reputation concerns: Daughety and Reinganum (2010)
- Law: D. Solove (2007, *San Diego Law Review*) "I've got nothing to hide"
- Strategic experimentation with Poisson bandits: Keller, Rady and Cripps (2005)
- Private learning in experimentation models:
  - Private payoffs but public actions: Rosenberg et al. (2007), Hopenhayn and Squintani (2011), Murto and Valimaki (2011) and Heidhues et al. (2015)
  - Public actions but private information arrival: Das and Klein (2020)
  - Partially observable actions: Guo and Roesler (2018)
  - R&D model with unobservable actions: Akcigit and Liu (2016)

Model

- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.

Model

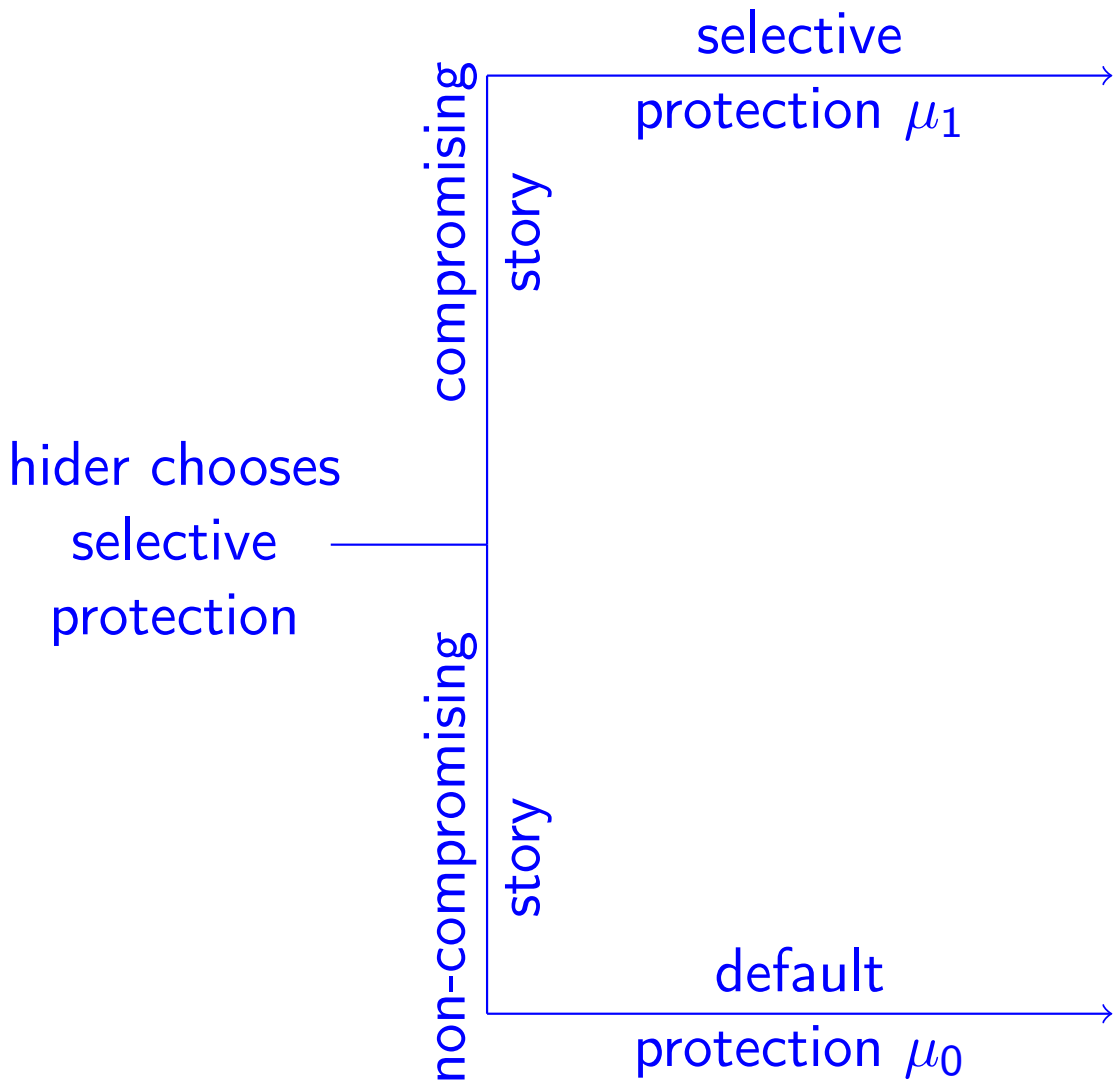
- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.
- The hider gets involved in a story of type $\theta \in \{0, 1\}$
  - $\theta = 1$: compromising story
  - $\theta = 0$: non-compromising story

  The seeker knows μ_0 and μ_1, and believes $\Pr(\theta = 1) = p$

Model

hider chooses
selective
protection

Model


selective
protection μ_1
compromising
story
hider chooses
selective
protection
non-compromising
story
default
protection μ_0

Model

- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.
- The hider gets involved in a story of type $\theta \in \{0, 1\}$
  - $\theta = 1$: compromising story
  - $\theta = 0$: non-compromising story

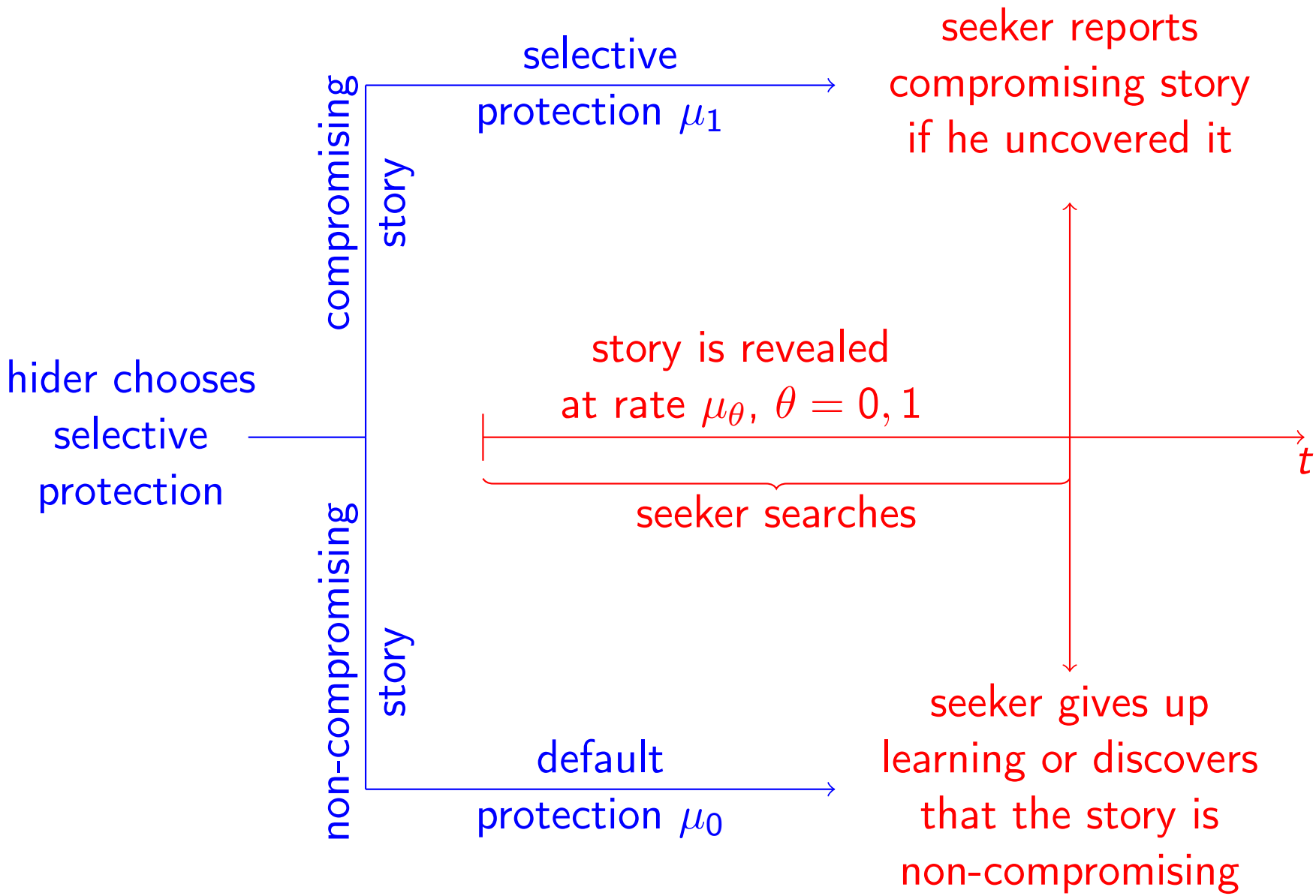
  The seeker knows μ_0 and μ_1, and believes $\Pr(\theta = 1) = p$
- The seeker learns the story through Poisson process with rate μ_θ and flow cost $c > 0$.

Model

- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.
- The hider gets involved in a story of type $\theta \in \{0, 1\}$
  - $\theta = 1$: compromising story
  - $\theta = 0$: non-compromising story

  The seeker knows μ_0 and μ_1, and believes $\Pr(\theta = 1) = p$
- The seeker learns the story through Poisson process with rate μ_θ and flow cost $c > 0$.
- The seeker gets 1 if he reports a compromising story and negative payoff if he reports a non-compromising story. To report the story, the seeker has to learn it.

Model


hider chooses
selective
protection
compromising
story
selective
protection μ_1
seeker reports
compromising story
if he uncovered it
story is revealed
at rate μ_θ, $\theta = 0, 1$
t
seeker searches
non-compromising
story
default
protection μ_0
seeker gives up
learning or discovers
that the story is
non-compromising

Model

- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.
- The hider gets involved in a story of type $\theta \in \{0, 1\}$
  - $\theta = 1$: compromising story
  - $\theta = 0$: non-compromising story

  The seeker knows μ_0 and μ_1, and believes $\Pr(\theta = 1) = p$
- The seeker learns the story through Poisson process with rate μ_θ and flow cost $c > 0$.
- The seeker gets 1 if he reports a compromising story and negative payoff if he reports a non-compromising story. To report the story, the seeker has to learn it.
- The hider minimizes Pr the seeker reports the story.
  - $\Rightarrow$ cost of protection $= 0$

Model

- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.
- The hider gets involved in a story of type $\theta \in \{0, 1\}$
  - $\theta = 1$: compromising story
  - $\theta = 0$: non-compromising story

  The seeker knows μ_0 and μ_1, and believes $\Pr(\theta = 1) = p$
- The seeker learns the story through Poisson process with rate μ_θ and flow cost $c > 0$.
- The seeker gets 1 if he reports a compromising story and negative payoff if he reports a non-compromising story. To report the story, the seeker has to learn it.
- The hider minimizes Pr the seeker reports the story.
  - $\Rightarrow$ cost of protection $= 0$
- No discounting

Model

- The hider commits to the level of selective protection, parametrized by $\mu_1 > 0$. The level of default protection $\mu_0 > 0$ is exogenous.
- The hider gets involved in a story of type $\theta \in \{0, 1\}$
  - $\theta = 1$: compromising story
  - $\theta = 0$: non-compromising story

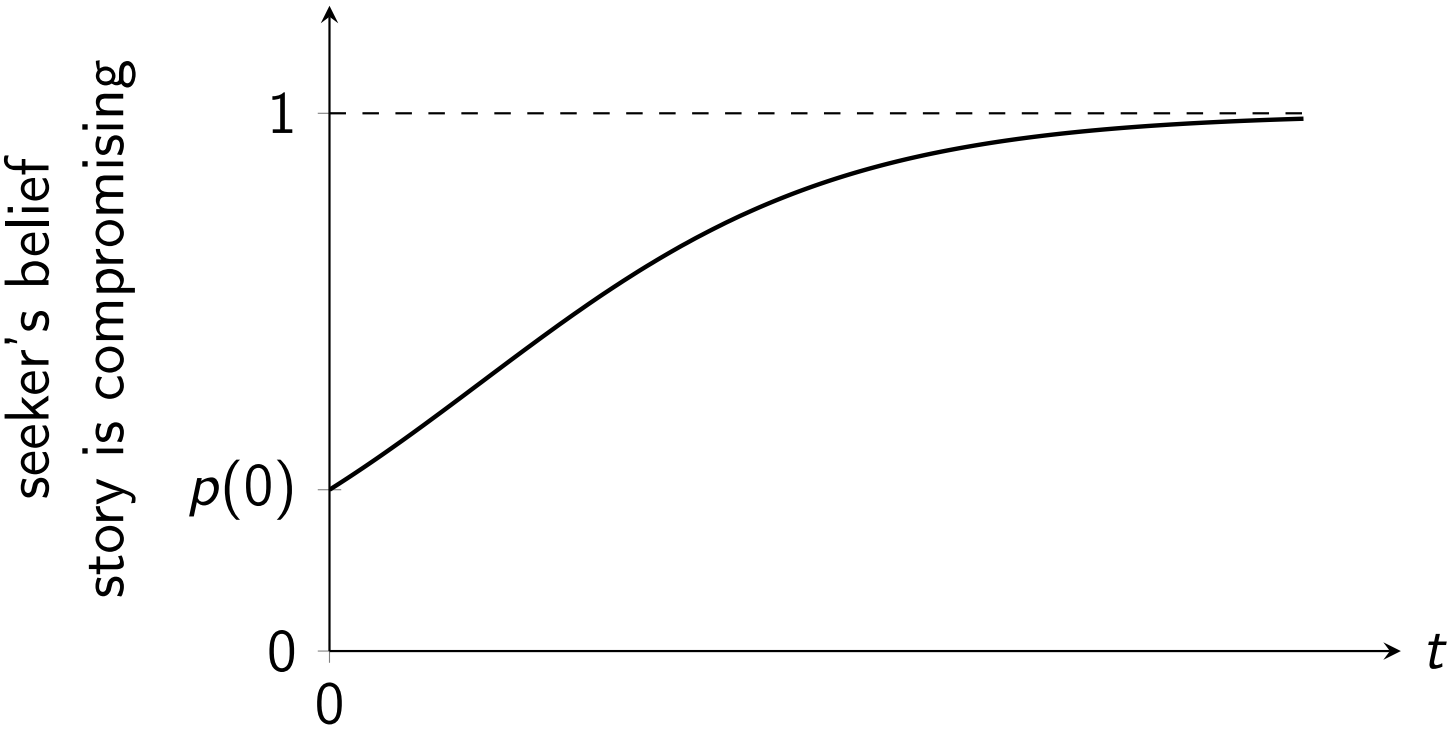
  The seeker knows μ_0 and μ_1, and believes $\Pr(\theta = 1) = p$
- The seeker learns the story through Poisson process with rate μ_θ and flow cost $c > 0$.
- The seeker gets 1 if he reports a compromising story and negative payoff if he reports a non-compromising story. To report the story, the seeker has to learn it.
- The hider minimizes Pr the seeker reports the story.
  - $\Rightarrow$ cost of protection $= 0$
- No discounting

Assumption

The hider can choose any μ_1 such that $\mu_1 \geq c/p$

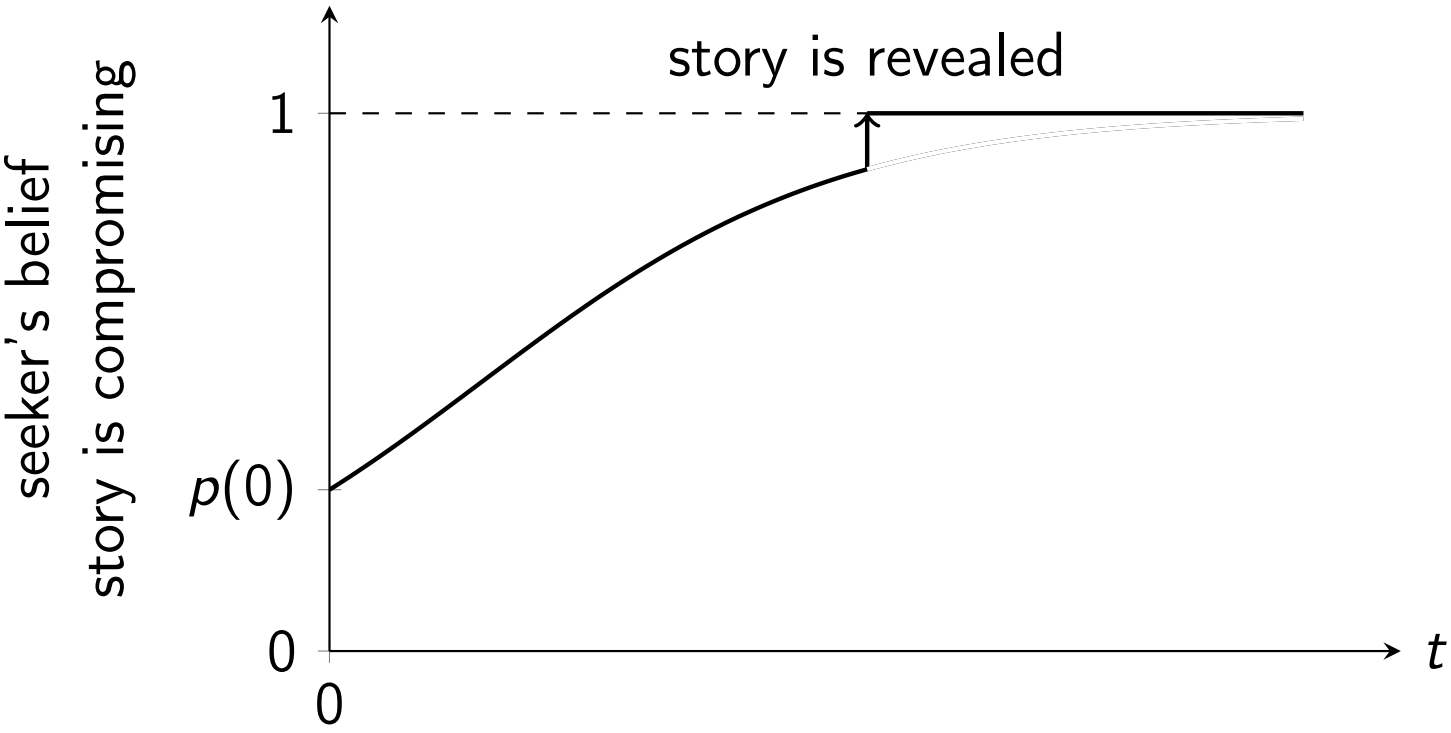
Results

If $c/p \leq \mu_1 < \mu_0$:



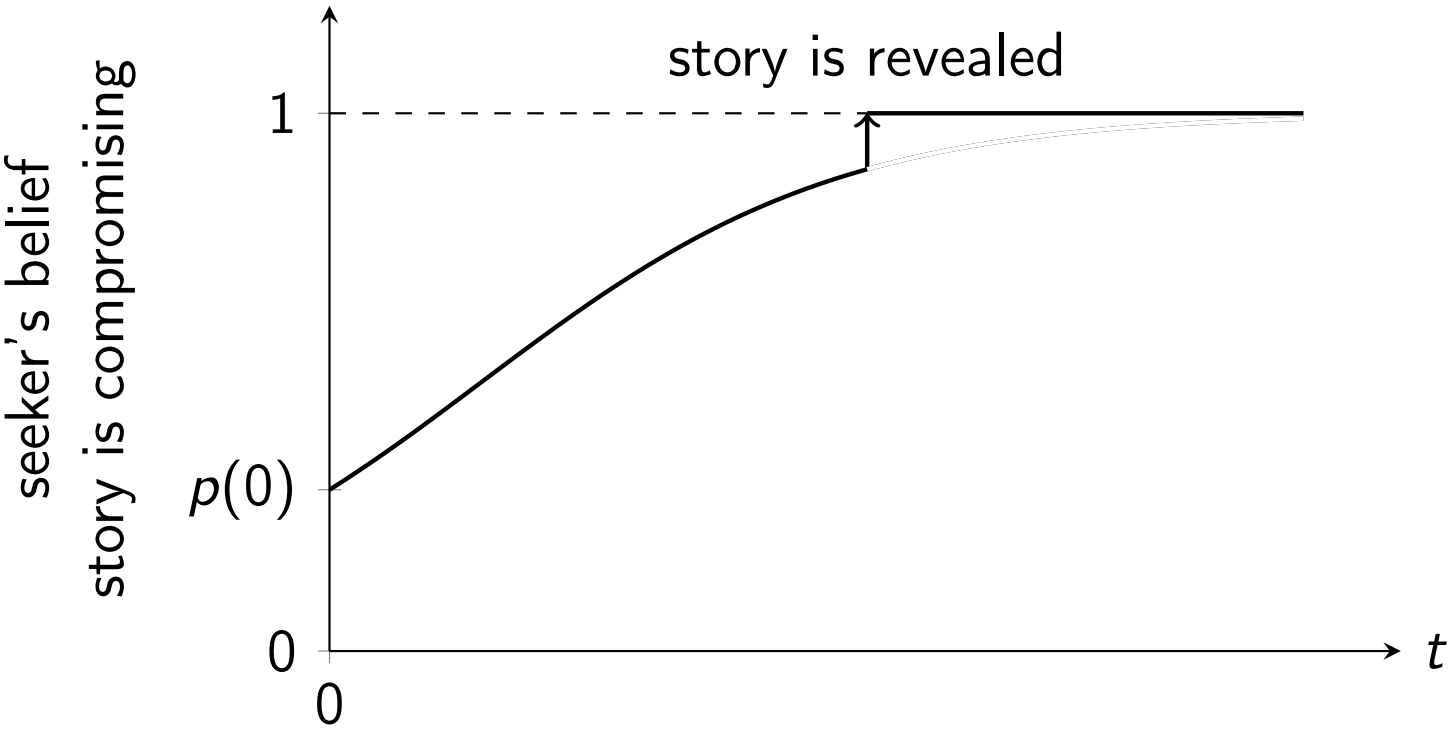
Results

If $c/p \leq \mu_1 < \mu_0$:



Results

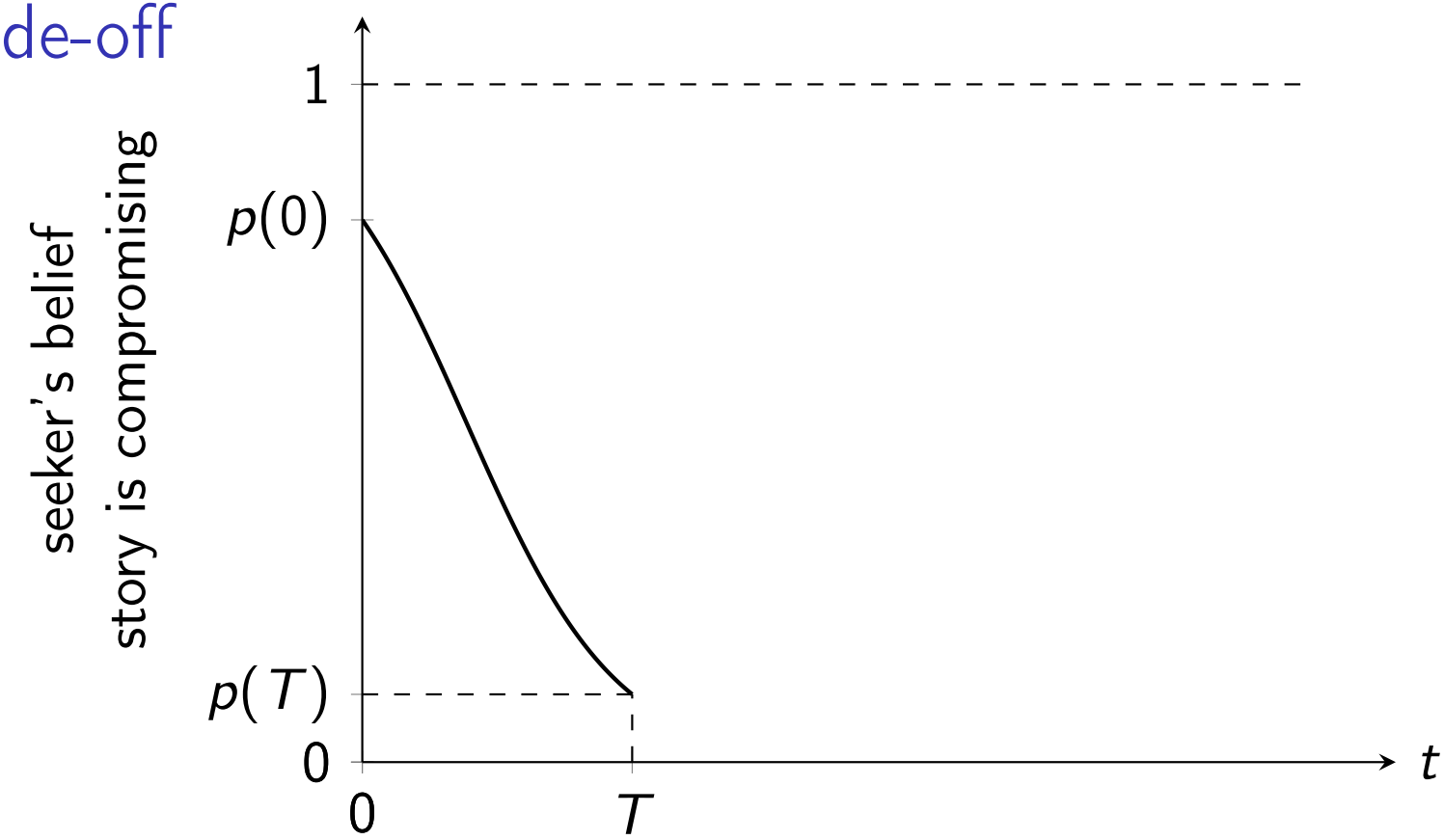
If $c/p \leq \mu_1 < \mu_0$:



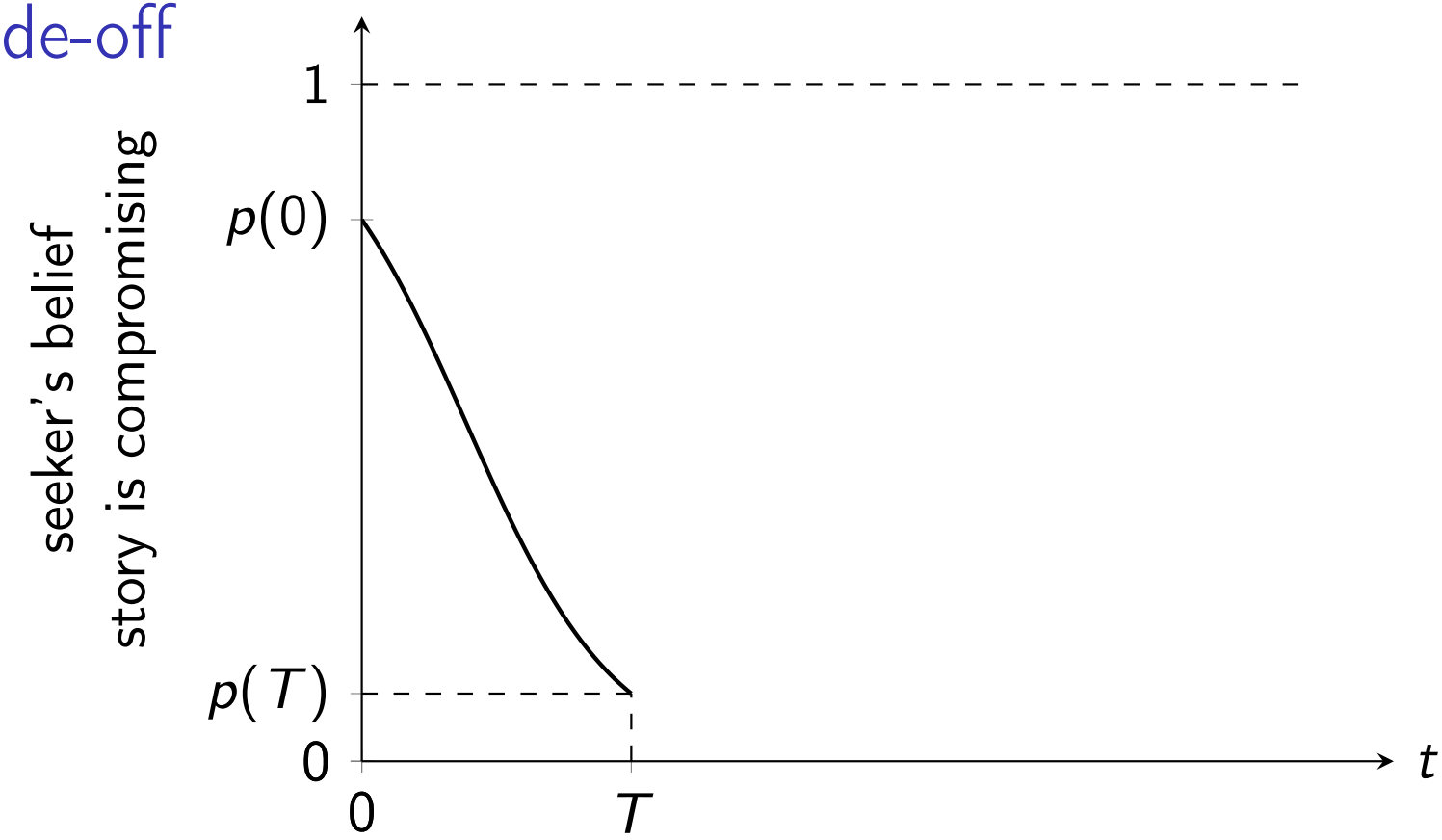
Main result

If $\mu_0 > c/p$ (default protection is weak), then the hider will not choose the strongest feasible selective protection: $\mu_1 \geq \mu_0$ at the optimum.

Trade-off

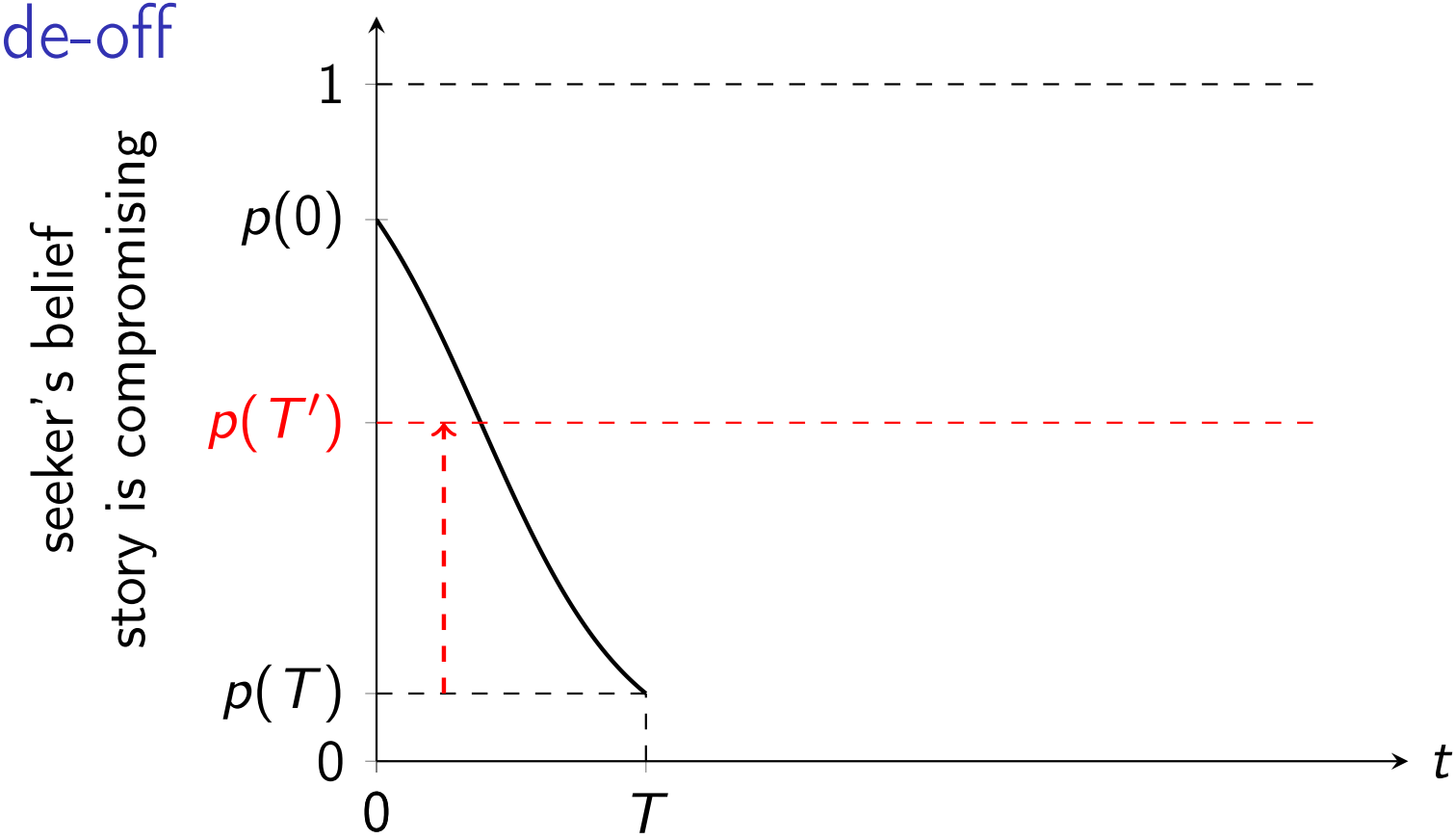


Trade-off


1
$p(0)$
$p(T)$
0
0
T
t
seeker's belief
story is compromising

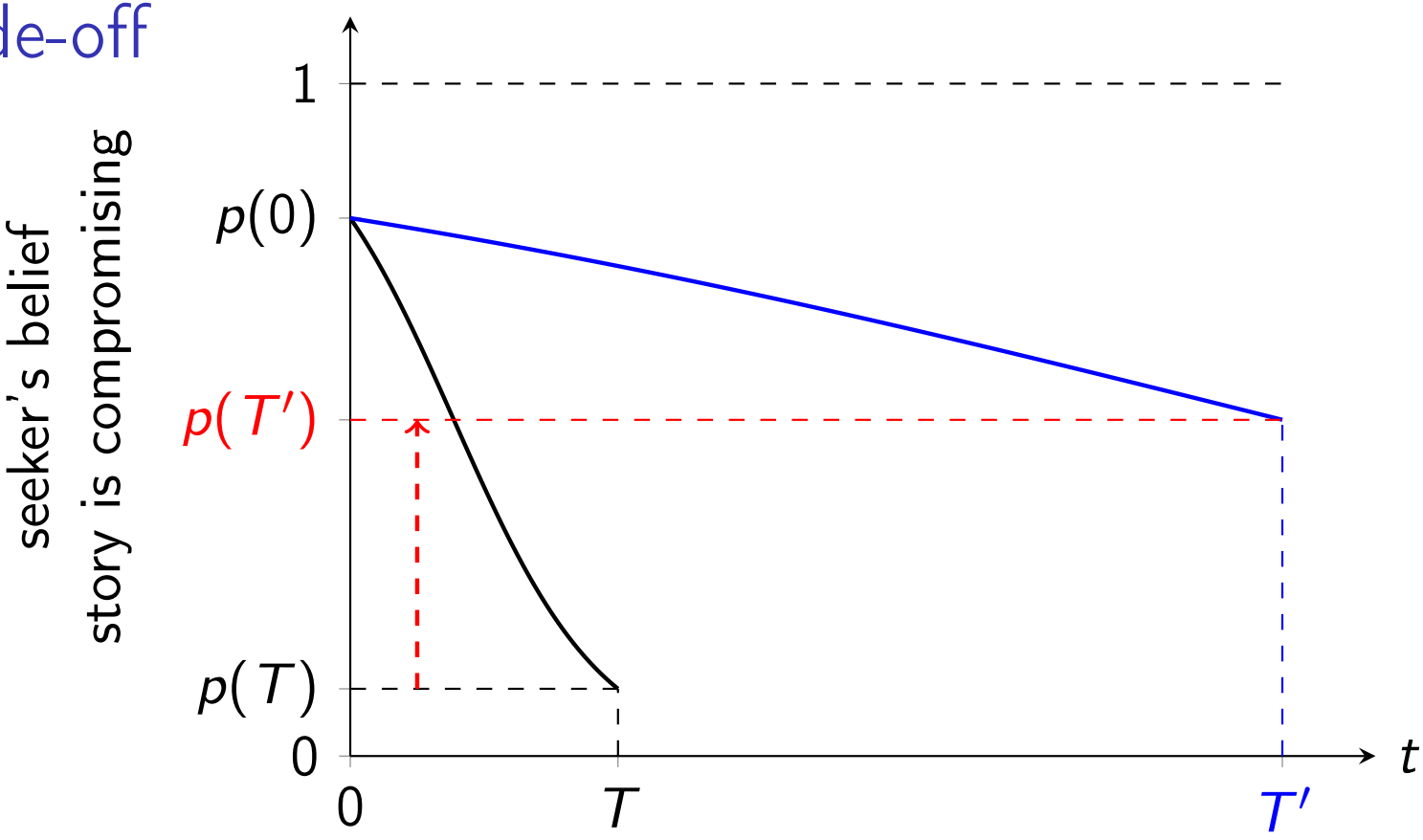

Pr(save reputation)
selective protection

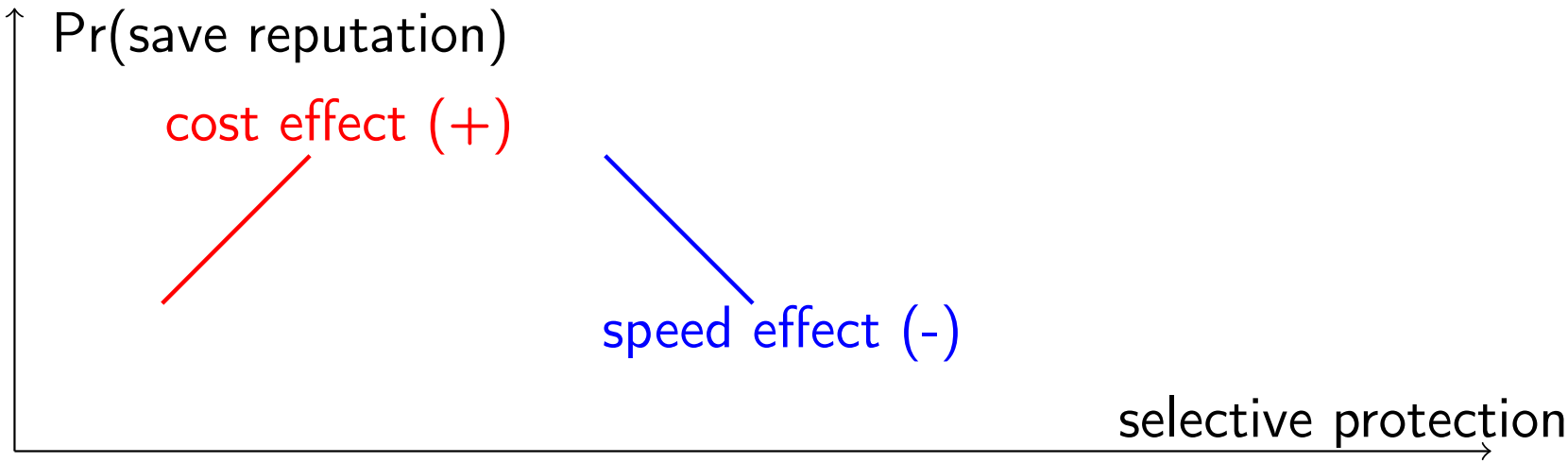
Trade-off


1
$p(0)$
$p(T')$
$p(T)$
0
0
T
t
seeker's belief
story is compromising

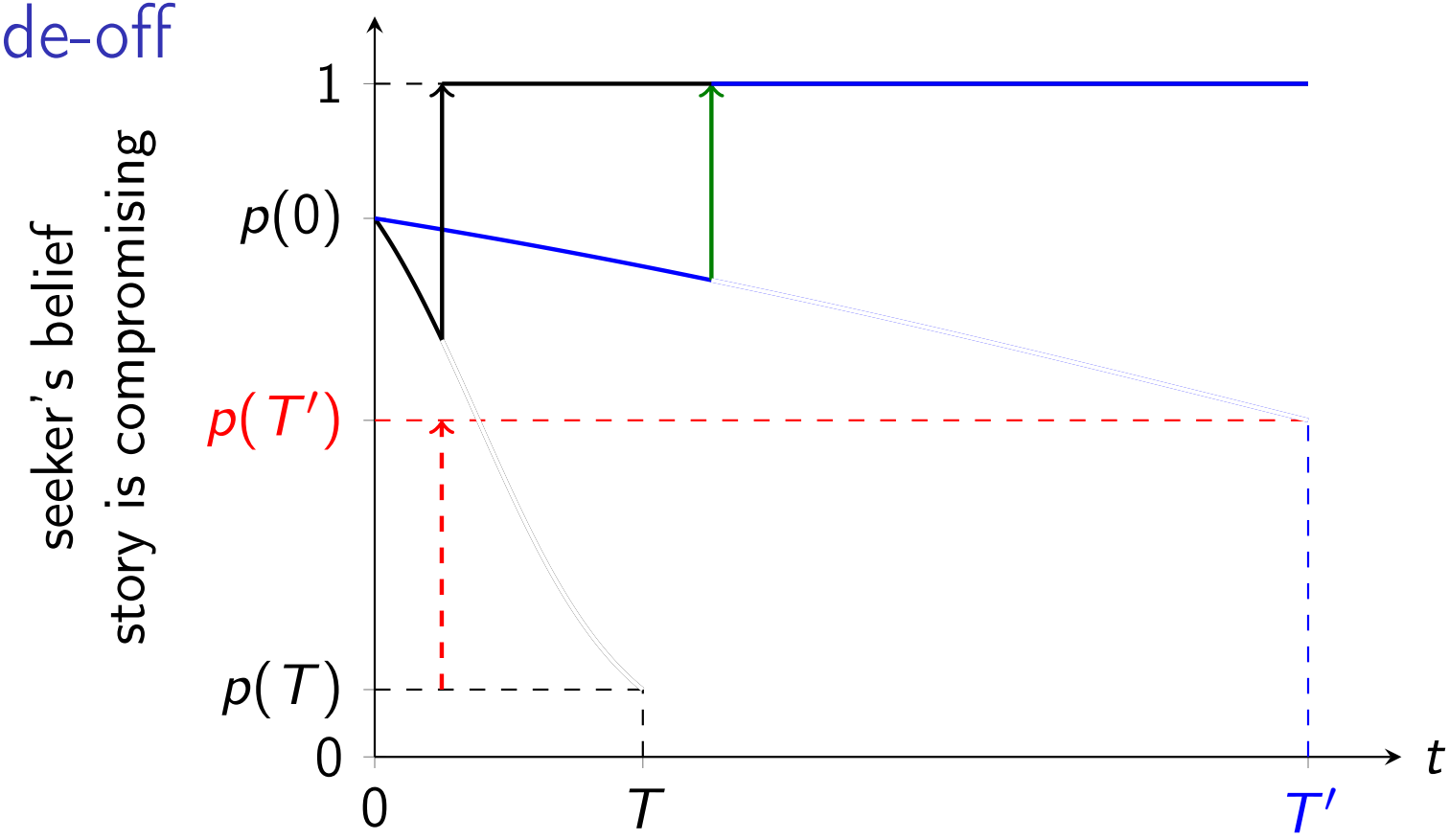

Pr(save reputation)
cost effect (+)
selective protection

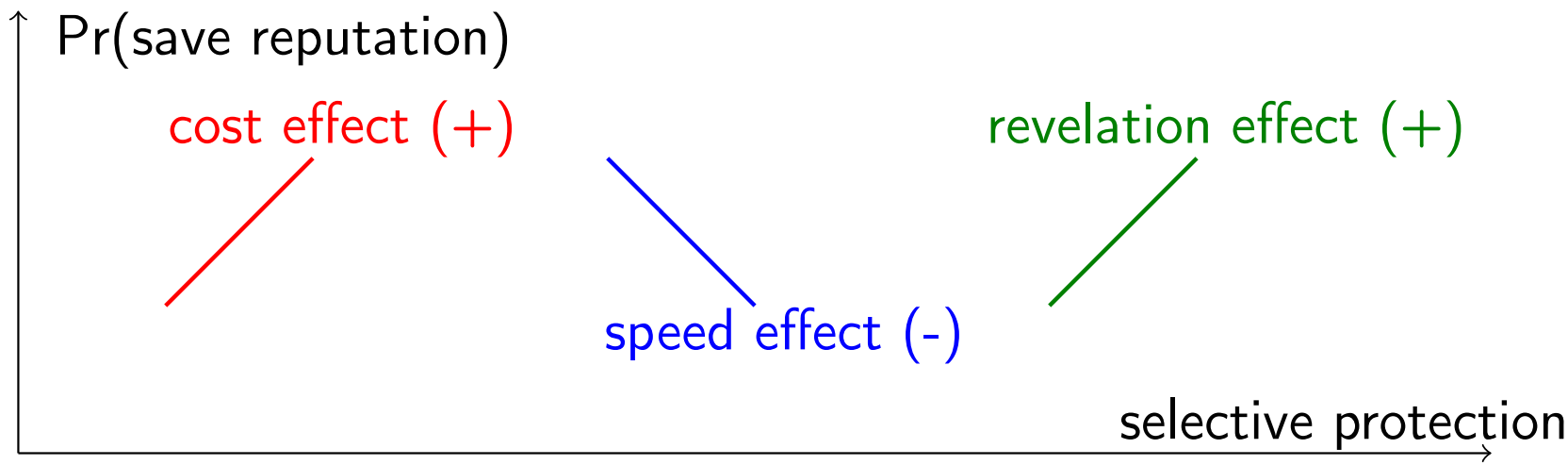
Trade-off


1
$p(0)$
$p(T')$
$p(T)$
0
seeker's belief
story is compromising
0
T
T'
t

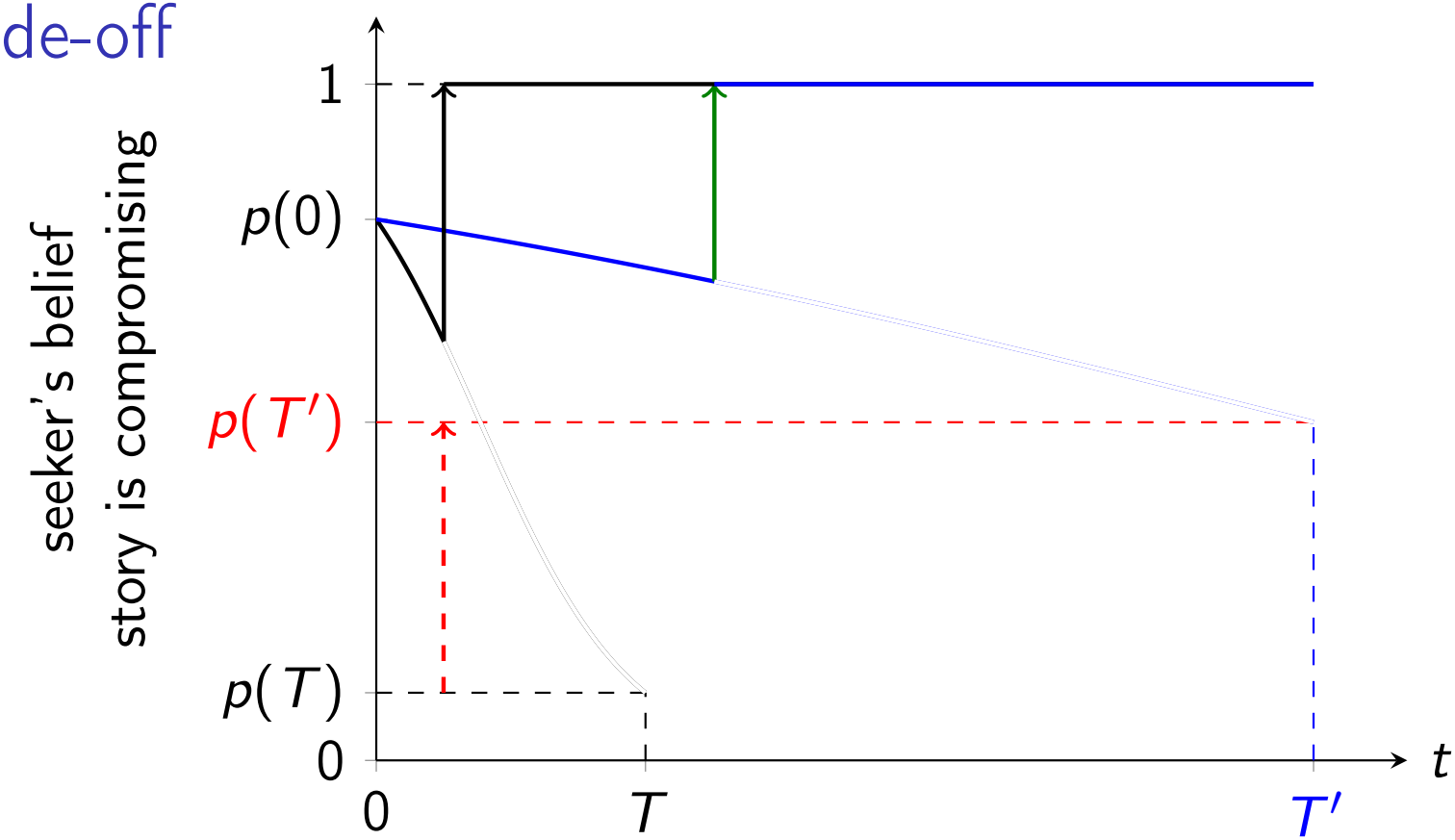

Pr(save reputation)
cost effect (+)
speed effect (-)
selective protection

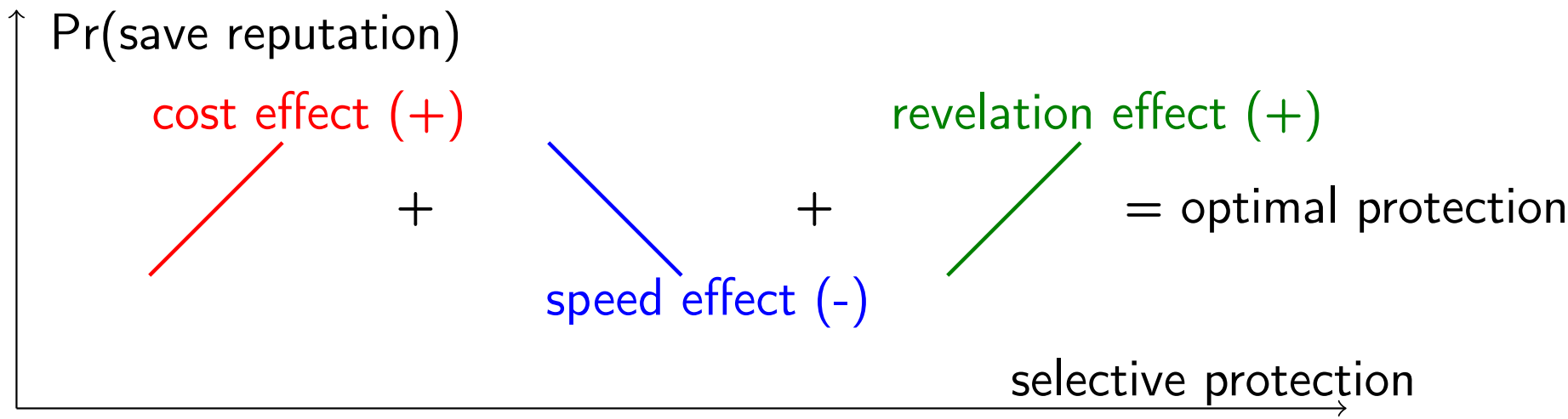
Trade-off


1
$p(0)$
$p(T')$
$p(T)$
0
0
T
T'
t
seeker's belief
story is compromising


Pr(save reputation)
cost effect (+)
revelation effect (+)
speed effect (-)
selective protection

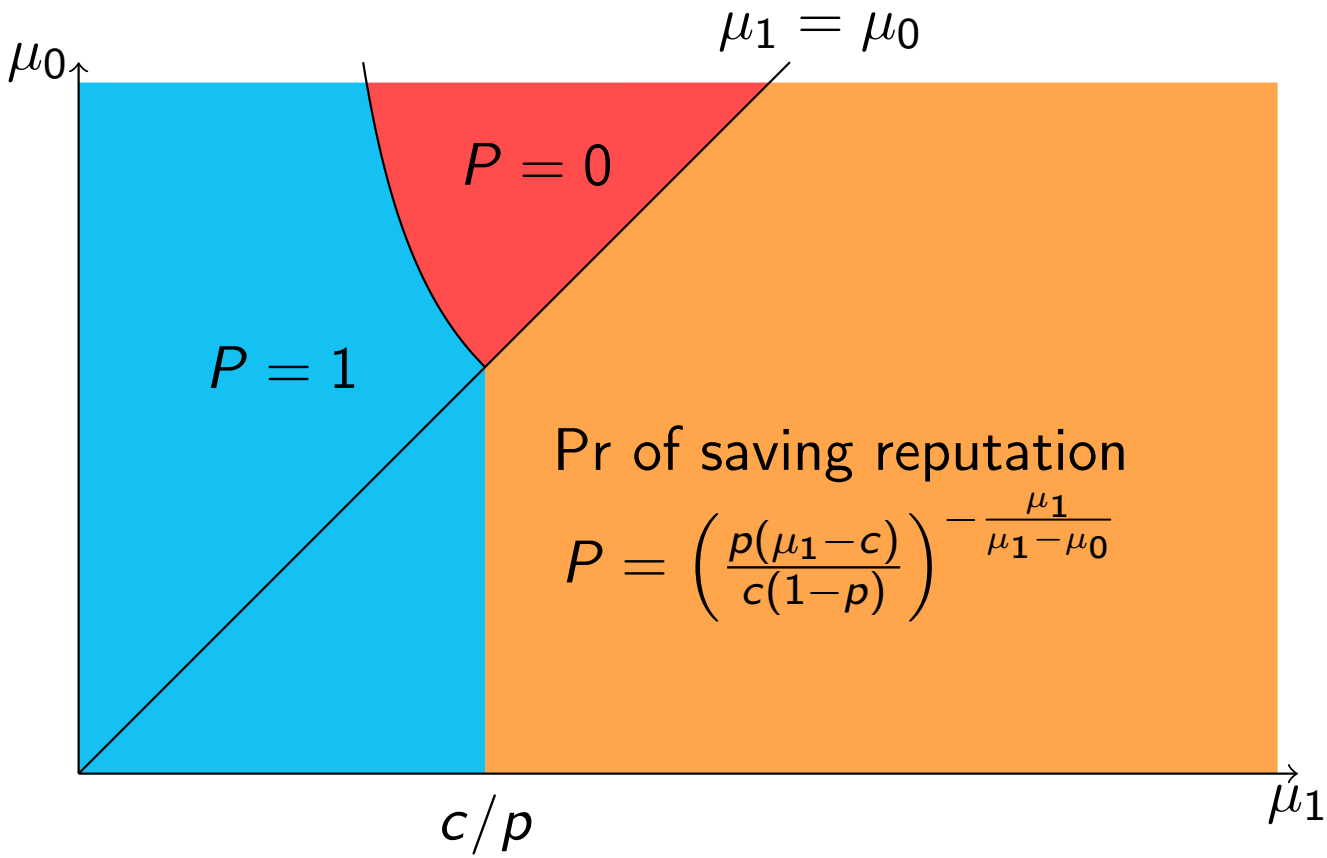
Trade-off


1
$p(0)$
$p(T')$
$p(T)$
0
seeker's belief
story is compromising
0
T
T'
t


Pr(save reputation)
cost effect (+)
+
speed effect (-)
+
revelation effect (+)
= optimal protection
selective protection

Optimal Protection

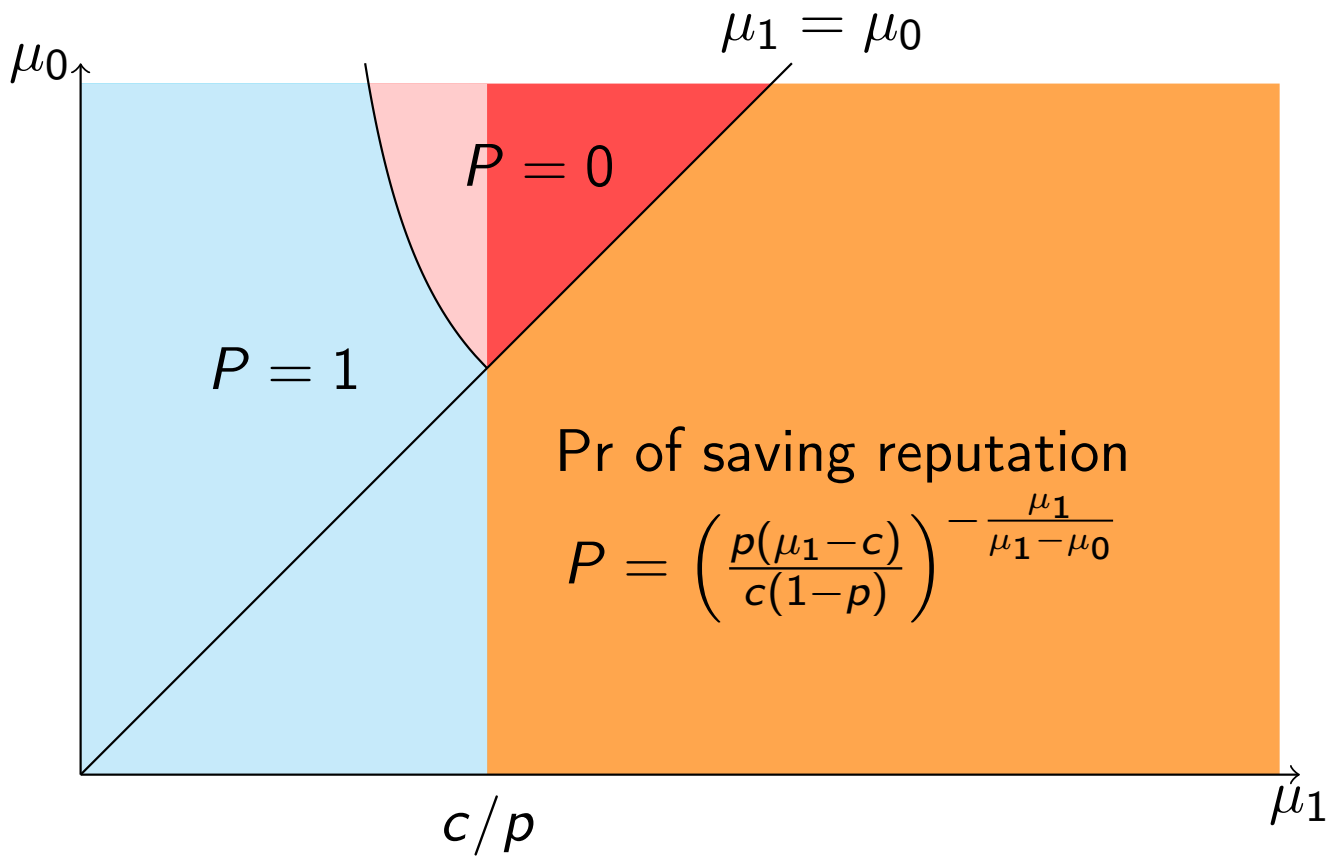
Theorem



Optimal Protection

Theorem

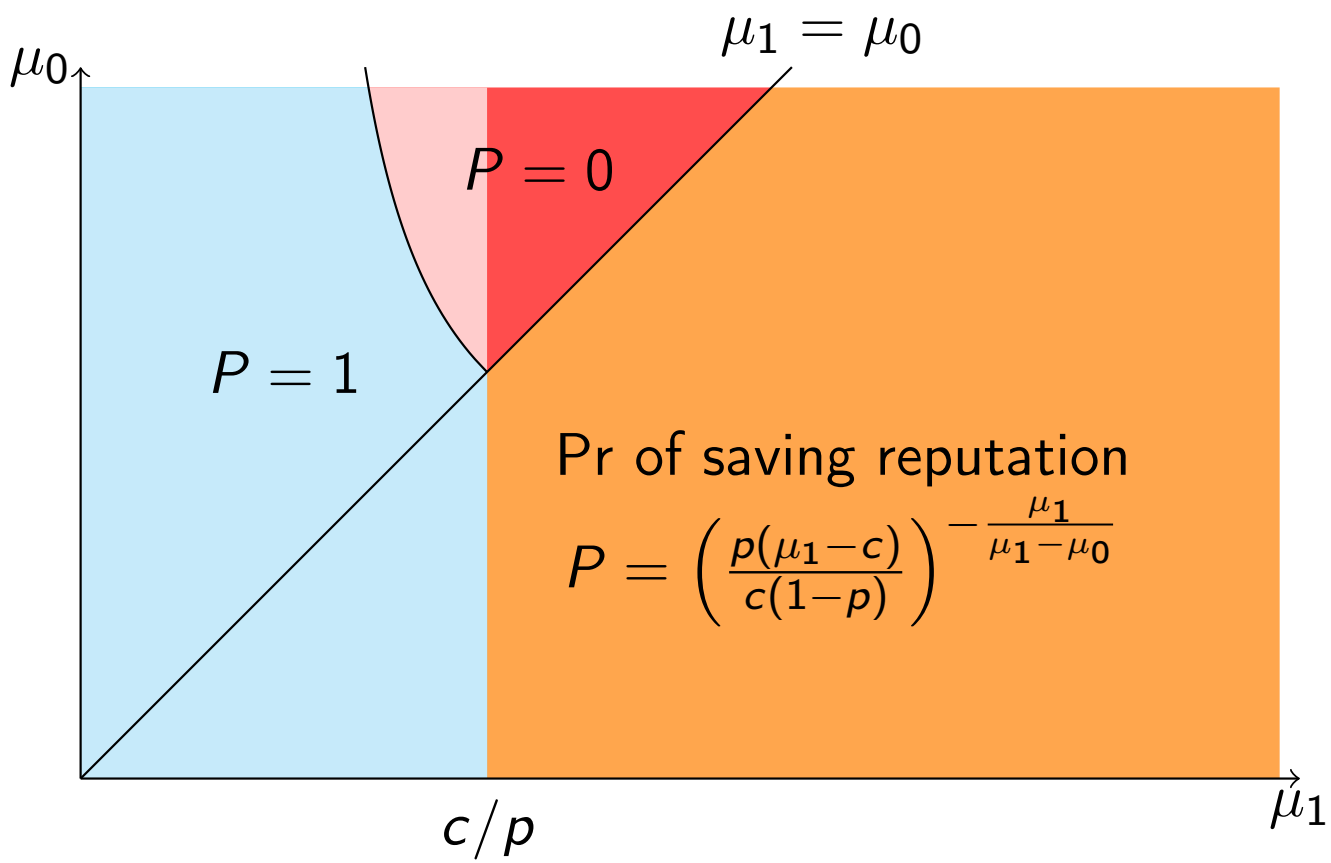
Under assumption $\mu_1 \geq c/p$,



Optimal Protection

Theorem

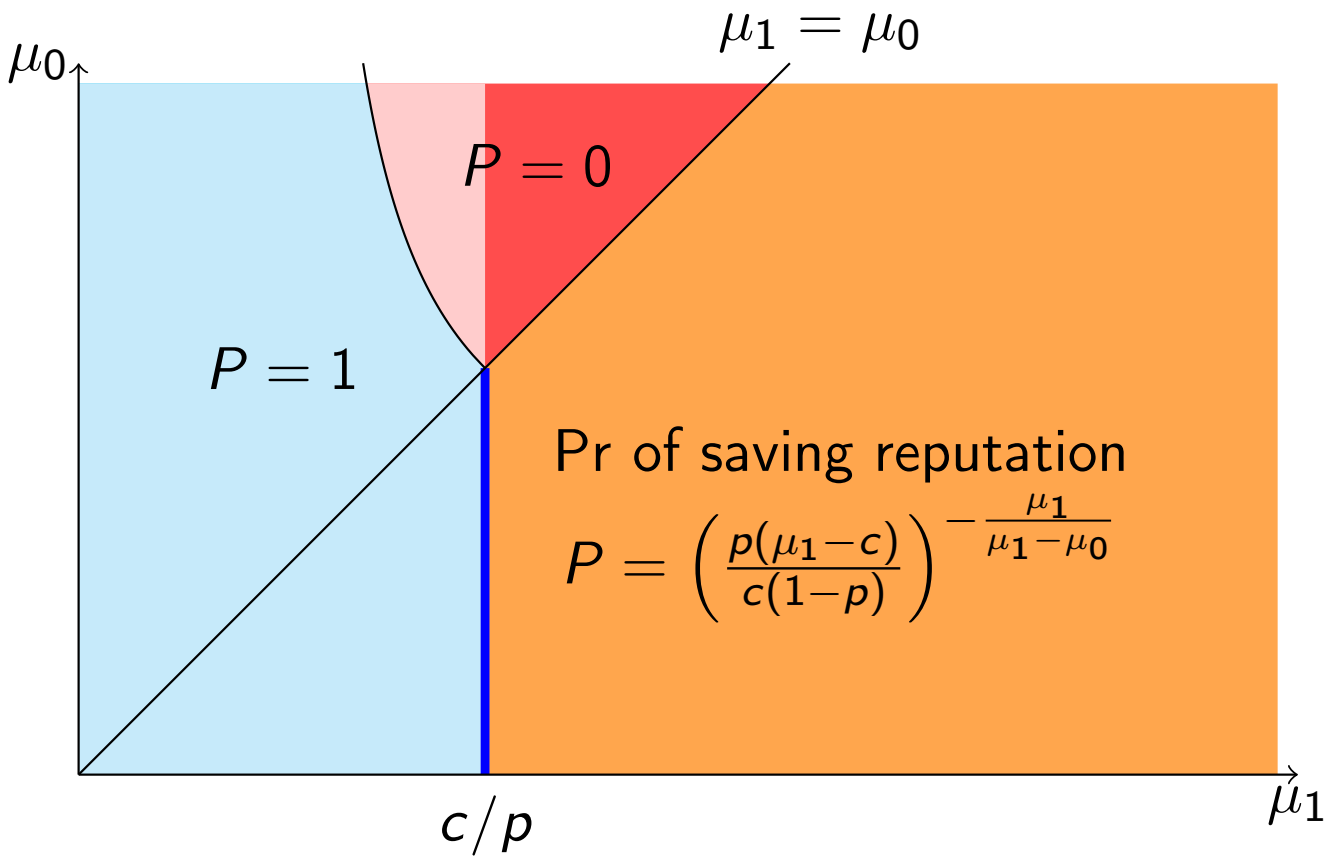
Under assumption $\mu_1 \geq c/p$, the optimal selective protection is characterized by μ_1^*, which is a function of μ_0, c, and p.



Optimal Protection

Theorem

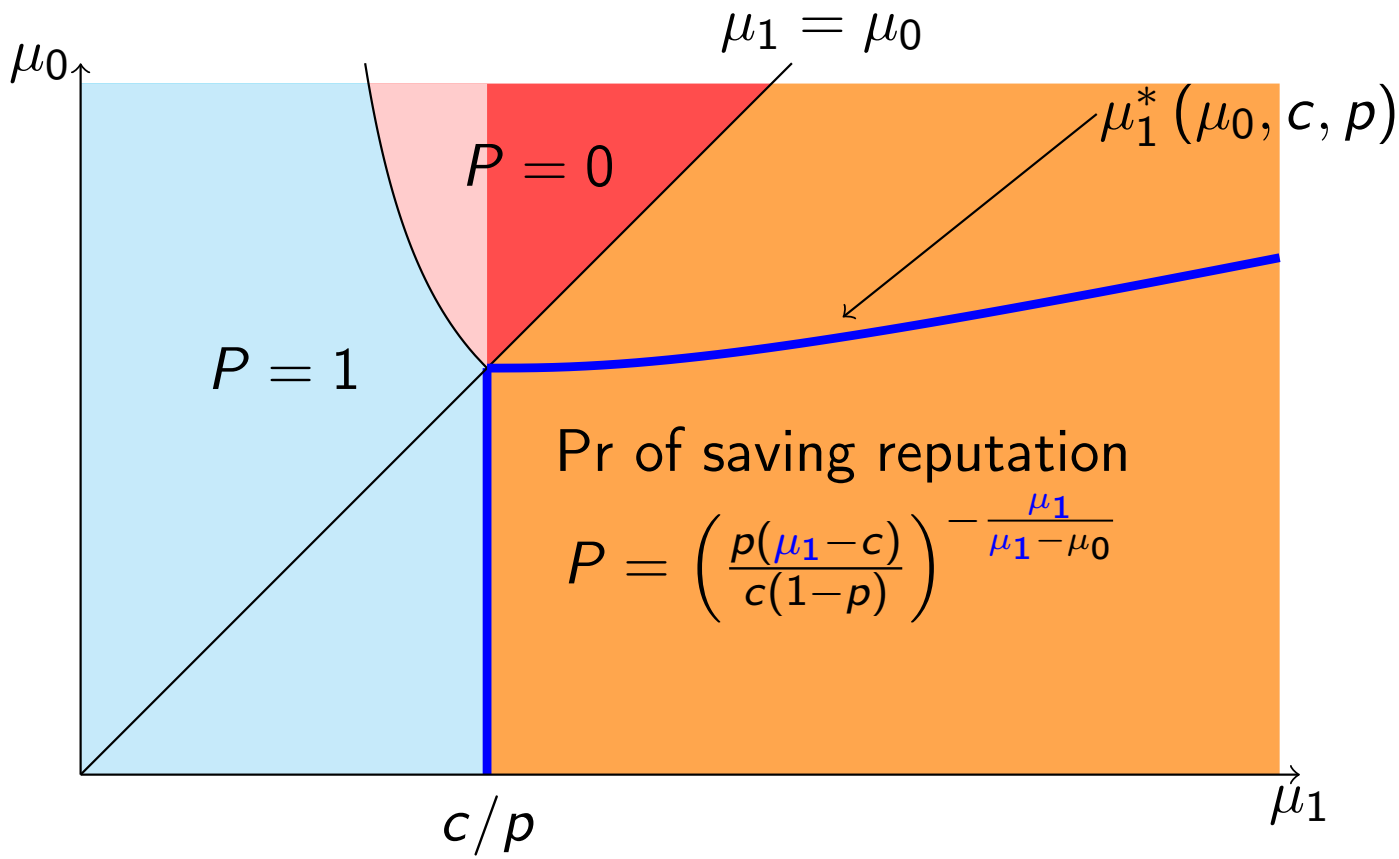
If $\mu_0 \leq c/p$, then $\mu_1^*(\mu_0, c, p) = c/p$.



Optimal Protection

Theorem

If $\mu_0 \leq c/p$, then $\mu_1^*(\mu_0, c, p) = c/p$. If $\mu_0 > c/p$, then $\mu_1^*(\mu_0, c, p) > \mu_0$ uniquely solves $\ln\left(\frac{p(\mu_1 - c)}{c(1-p)}\right) - \frac{\mu_1(\mu_1 - \mu_0)}{\mu_0(\mu_1 - c)} = 0$.



Comparative Statics

$\mu_1^*(\mu_0, c, p)$ balances *Speed* vs *Duration* of learning

Comparative Statics

$\mu_1^*(\mu_0, c, p)$ balances *Speed* vs *Duration* of learning

Prior Belief

$\mu_1^*(\mu_0, c, p)$ is increasing in p.

- $\uparrow p \Rightarrow$ learning interval is wider $\Rightarrow \uparrow$ learning duration $\Rightarrow \uparrow \mu_1$ to increase speed and lower the duration

Comparative Statics

$\mu_1^*(\mu_0, c, p)$ balances *Speed* vs *Duration* of learning

Prior Belief

$\mu_1^*(\mu_0, c, p)$ is increasing in p.

- $\uparrow p \Rightarrow$ learning interval is wider $\Rightarrow \uparrow$ learning duration $\Rightarrow \uparrow \mu_1$ to increase speed and lower the duration

Cost

$\mu_1^*(\mu_0, c, p)$ is decreasing in c.

- $\downarrow c \Rightarrow$ learning stops later $\Rightarrow \uparrow$ learning duration $\Rightarrow \uparrow \mu_1$ to increase speed and lower the duration

Comparative Statics

$\mu_1^*(\mu_0, c, p)$ balances *Speed* vs *Duration* of learning

Prior Belief

$\mu_1^*(\mu_0, c, p)$ is increasing in p.

- $\uparrow p \Rightarrow$ learning interval is wider $\Rightarrow \uparrow$ learning duration $\Rightarrow \uparrow \mu_1$ to increase speed and lower the duration

Cost

$\mu_1^*(\mu_0, c, p)$ is decreasing in c.

- $\downarrow c \Rightarrow$ learning stops later $\Rightarrow \uparrow$ learning duration $\Rightarrow \uparrow \mu_1$ to increase speed and lower the duration

Default Protection

$\mu_1^*(\mu_0, c, p)$ is increasing in μ_0.

- $\uparrow \mu_0 \Rightarrow \downarrow$ speed of learning $\Rightarrow \uparrow \mu_1$ to increase speed
- $\Rightarrow$ general and selective types of protection are *complements*

Privacy Debate

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

Goal protection of sensitive information

Problem Default 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

Question Could providing tools for selective protection be a solution?

Privacy Debate

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

Goal protection of sensitive information

Problem Default 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

Question Could providing tools for selective protection be a solution?

Answer No! Tools that facilitate selective protection might not be used in practice in the absence of good default protection.

Extension: Many Seekers

- Suppose the hider cannot control μ_1 but can choose the number of seekers $n \geq 1$.
- Private learning: the seekers do not communicate with each other
- Competition: only the first seeker to report a compromising story derives a positive payoff (reports are public)

Extension: Many Seekers

- Suppose the hider cannot control μ_1 but can choose the number of seekers $n \geq 1$.
- Private learning: the seekers do not communicate with each other
- Competition: only the first seeker to report a compromising story derives a positive payoff (reports are public)

Revelation effect $\uparrow n \Rightarrow \uparrow$ Pr story is revealed when the seekers are searching for it

Speed effect $\uparrow n \Rightarrow p(t) \downarrow$ faster

No cost effect n does not change the stopping belief threshold $p(T)$

Extension: Many Seekers

- Suppose the hider cannot control μ_1 but can choose the number of seekers $n \geq 1$.
- Private learning: the seekers do not communicate with each other
- Competition: only the first seeker to report a compromising story derives a positive payoff (reports are public)

Revelation effect $\uparrow n \Rightarrow \uparrow$ Pr story is revealed when the seekers are searching for it

Speed effect $\uparrow n \Rightarrow p(t) \downarrow$ faster

No cost effect n does not change the stopping belief threshold $p(T)$

Theorem

Under assumption $\mu_1 \geq c/p$, the optimal privacy protection is $n = +\infty$, that is, open access is optimal.

- Without discounting, speed effect always gets an upper hand over revelation effect
- With discounting, bang-bang solution

Reality check

Reality check

The Anti-Corruption Foundation

The ACF exposes corruption cases on the basis of publicly available data.

Reality check

The Anti-Corruption Foundation

The ACF exposes corruption cases on the basis of publicly available data.

Fall after news

A flurry of news about a company often follows by a fall in the stock price.