Bayesian Persuasion as Information Design: Focus, Methods and Insights

Preliminary Version

Renjie Zhong

2021 PPE
RUC

May 2023

Outline

Introduction

Driving Forces of Micro Behavioral

- The micro-behavior of an agent depends on his beliefs μ_i, his feasible choices A_i, the resulting final payoffs u_i, neighbors G_i and some idiosyncratic constraints.
- Design Problem: Designer design the game structure to implement/realize optimal/revenue-maximizing outcome

Design Problem

- Mechanism/Market/Network Design as Institution/Organization Design
  1. mechanism design with "monetary" incentives (transferable or non-transferable utility): steer the agent(s) decisions by changing their payoff consequences
  2. delegation/redistribution policy deisgn: steer the agent(s) decisions by constraining the set of feasible actions
  3. matching/market design without "money": steer the agent(s) decisions by designing the rules whereby reports about preferences map to final allocations of objects
  4. network intervention/design: steer the agent(s) decisions by constraining the set of players/ changing their payoff consequences

Information Design: Motivation

- An agent's beliefs are an important driver of his behavior and can be influenced by information transmission from another agent, motivating the problem of **information design**
- In information design, payoff functions and feasible outcomes (i.e., the game) taken as given
- object of design: information of the agent(s)—hence, the beliefs driving choices
  1. different characteristics of information (public/private, hard/soft, ambiguous/certain)
  2. commitment/no commitment
  3. bayes rule

Information Design: Focus

- Desipite of different situations, we always concern these problems in information design:
  1. Feasibility: what is the scope for changing the agent's behavior by designing his information environment?

     the probability of successful persuasion
  2. Optimality: what is the optimal information for the agent from the viewpoint of its designer?

     Optimal Information Structure, cardinality of signals $|S|$, posterior distribution...
  3. Welfare: when persuasion is beneficial/detrimental to the sender/receiver?
  4. Robustness: (partially) informed receiver, witness, side-communication

Group Persuasion: Focus

- With multiple agents, we also care about the timing/sequence of the persuasion
- How setup affects the information releavation? (Comparative Static Analysis)
  1. the alignment/congruence of preference between senders and receivers/ within senders/receivers
  2. the number of receivers/senders

Bayesian Persuasion

- Bayesian Persuasion impose a critical assumption on the general information design: commitment
- We can intepretate it as a constraint on information structure – bayesian plausibility (martingale property/consistency/committment) under bayesian updating
- This subfield is recognized as the contribution of Kamenica and Gentzkow(2011).

  some debates (Aumann and Maschler,1995;Benot and Dubra,2011)

This Paper

- This Paper focuses on:
  1. **a comprehensive framework of the extensions** of basic game struture (N, A, T, Ω, H) (to my knowledge it is the first work)
  2. a very simple but comprehensive survey of **methods and perspectives**
  3. concavification method due to its geometry illustration of **strategic insights**

Some Remarks

- A preliminary survey open to criticism!
  1. pre 可能更偏重于呈现各类文章的主要结论而显得增添了很多图表例子 (数学符号能省则省了),survey 写作会更侧重归纳而不会像本次 pre 一样走马观花看很多图文,
  2. 本文选取的文献大多使用 concavification,其良好简洁的几何性质方便阐释文本的 insights,但文本尚未从数理上系统考察其 robustness。Mathevet et.al(2020) 提供了初步的思考,但是更多关注的 higher order beliefs 等传统的博弈议题的关联,本文会致力于在 survey 基础上去做更 general 的研究进行统摄
  3. 漏洞: 没有统摄好各类 extension 下 focus 问题的 insights 如何变化
  4. 由于时间和精力关系,pre 涉及的约十余篇文章全是 top5 的文献,且以纯理论居多,这一领域近年来已有大量偏纯理论的精干作品涌现,本文难免挂一漏万
- I will only briefly review it to spare time for extension section!

Related Surveys and Notes

- Kamenica (2019; 2022): concavification, its extensions (multiple players and dynamics) and leading economic examples
- Bergemann and Morris (2019): a distinction between literal/metaphorical information design
  1. literal: optimal choice of information structure
  2. metaphorical: optimal (action recommendation) mechanism under different information structures
- Bergemann and Bonatti (2019): a framework of information selling
- Lecture note/slides:
  1. Introductory slides and focusing on BCE and Concavification: Morris-Bonn Lectures(2018),Sandomirskiy(2020),Starkov(2022)
  2. A systematical exploration and focusing on the number of receivers: Galerpti(2022)

Benchmark Model

Benchmark Model: Setup

- Players: **one** sender S/receiver R with different priors[1] μ_0^S, μ_0^R
- Notations: $\omega \in \Omega$: state space; $v(a,\omega), u(a,\omega)$: sender,receiver′s payoff
- Action[2] Space: $\pi : \Omega \to \Delta S$ (S: the set of signal realizations)
  1. zero marginal/common fixed cost of signals
  2. all information structures are feasible
  3. public signals
- Updating Rules: **Bayesian Updating** $\mu^S = \mu_s^B(\omega; \mu_0, \pi) = \frac{\pi(s|\omega)\mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')\mu_0(\omega')}$

[1] we take the fashion of (Alonso and Camara,2016a) without adding too much complexity to (Kamenica and Gentskov,2011)

[2] also called signal structure, information structure, experiment, Blackwell experiment, or data-generating process

Benchmark Model: Setup

- Game Rules:
  1. commitment power[3][4]: denote $\tau(\mu) = \sum_{\omega\in\Omega}\sum_{s:\mu_s=\mu}\pi(s \mid \omega)\mu_0(\omega)$, then $\sum_{\mu\in \text{ supp } \tau}\mu(\omega)\tau(\mu) = \mu_0(\omega), \quad \omega \in \Omega$
  2. Designer-preferred equilibrium: choose $a \in A^*(\mu) = \arg\max_{a\in A}\mathbb{E}_\mu[u(a,\omega)]$ that maximizes $\mathbb{E}_\mu[v(a,\omega)]$ when $|A^*(\mu)| \geq 2$
- Timeline: designer commits $\pi \Rightarrow \omega$ realizes $\Rightarrow$ agent observes s, updates her belief and chooses her action $\Rightarrow$ payoffs realized
- Statics: only one period
- We will relax at least one assumptions in extension and check its robustness

[3] all called bayesian plausibility/consistency/martingale property (especially in dynamic setting (Ely et al., 2015))

[4] compared to other info, we can interpret it as no signaling through info structure, signals with objective meaning, info transmission with reputation foundation (Best and Quigley(2017), Mathevet et al.(2019))

Concavification (KG,2011)

Sender′ Problem

Define $\hat{v}(\mu) = \mathbb{E}_\mu[v(\hat{a}(\mu), \omega)]$, then sender′ problem is:

$$v^* = \max_\tau \sum_{\mu \in \operatorname{supp} \tau} \hat{v}(\mu)\tau(\mu)$$
$$\text{s.t. } \sum_{\mu \in \operatorname{supp} \tau} \mu\tau(\mu) = \mu_0$$

- Concavification: $[\mathrm{CAV}(f)](\mu) = \sup\{z \mid (\mu, z) \in \mathrm{co}(f)\}$
- Considering $[\mathrm{CAV}(\hat{v})](\mu)$ and pick up the optimal signal![5]

[5] We need two additional assumptions to guarantee the convex combination, one is the willingness to share $(\exists \mu, \hat{v}(\mu) > \mathbb{E}_\mu\left[v\left(\hat{a}\left(\mu_0\right), \omega\right)\right])$, another is a technical assumption called ”local continuity” at μ_0 $(\exists \varepsilon > 0 s.t. \mathbb{E}_\mu[u(\hat{a}(\mu), \omega)] > \mathbb{E}_\mu[u(a, \omega)] + \varepsilon, \quad \forall a \neq \hat{a}(\mu))$

Concavification: A Graphical Illustration

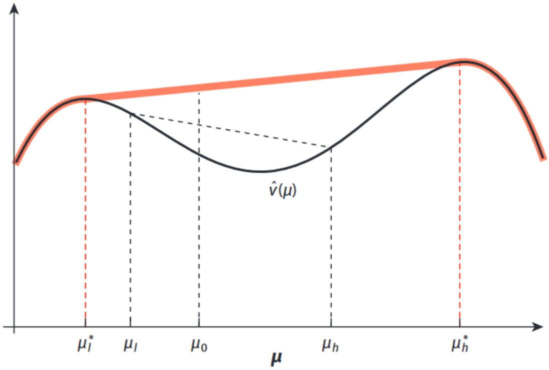


Figure 1

Sender's value function and its concavification (*thick red line*).

Pros and Cons of Concavification

- pros: always nice interpretation and allows for convex analysis; robustness to the analysis of various extensions
- cons: difficulty to characterize/derive the optimal signals; high dimensions/multiple players worsens it
- Other approaches:
  1. Myersonian Approach (Bergemann and Morris,2016; Morris,2019): transform the seletion of optimal signals into the selection of optimal mechanism under different information structures
  2. Rothschild-Stiglitz Approach (Gentzkow and Kamenica,2016): a special class with uncountable state spaces and Sender's payoff depends only on the mean of Receiver's posterior.
  3. computational methods (Dughmi and Xu,2016; Dughmi, 2017)

Myersonian Approach and Duality

- Bergemann and Morris (2016,2018) transform it into a linear programming problem
- each $\pi(\cdot \mid \omega)$ induces a distribution $x(\cdot \mid \omega) \in \Delta(A)$ over actions:

$$x(a \mid \omega) = \sum_{s:\hat{a}(\mu_s)=a} \pi(s \mid \omega)$$

Primal problem

choose $x \in \mathbb{R}^{A\times\Omega}$ to solve:

$$\max \mathcal{V}(x) = \sum_{\omega\in\Omega, a\in A} v(a,\omega)x(a \mid \omega)\mu_0(\omega) \quad \text{s.t.}$$

1. (O) Obedience: $\sum_{\omega\in\Omega} [u(a,\omega) - u(a',\omega)]\, x(a \mid \omega)\mu_0(\omega) \geq 0$ for all $a, a' \in A$
2. (C) Consistency: $\sum_{a\in A} x(a \mid \omega) = 1$ for all $\omega \in \Omega$
3. (NN) Non-negativity: $x(a \mid \omega) \geq 0$ for all $(a,\omega) \in A \times \Omega$

Remarks

- Remarks:
  1. equil selection matters!
  2. a dual problme (linear programming)
  3. applications in pricing theory
- A very substantial literature strand: Bergemann-Morris (2016), Kolotilin (2018), Galperti and Perego (2018), Morris, Oyama and Takahashi(2023)

Rothschild-Stiglitz and Walrasian Economy Perspective

- Gentzkow and Kamenica (2016) proposes a way to tackle a special class with uncountable state spaces and Sender's payoff depends only on the mean of Receiver's posterior.
- Dworczak and Martini (2018) proposes a price-theoretic approach to Bayesian persuasion by establishing an analogy between the Sender's problem and finding Walrasian equilibria of a Persuasion Economy
- A very substantial literature strand: firm-consumer pricing problem (Roesler and Szentes,2017; Condorelli and Szentes,2020)

Extensions

Extensions

- We consider the extensions of benchmark models by relaxing these assumptions/setups:
  1. players: single sender/receiver $\Rightarrow$ multiple senders/receivers
  2. action space: costless signals $\Rightarrow$ costly signal sending/processing
  3. updating rules: bayesian updating $\Rightarrow$ non-bayesian updating
  4. game rules:
     1. commitment power $\Rightarrow$ partial commitment
     2. designer-preferred equilibrium selection $\Rightarrow$ designer-worst/general equilibrium selection
  5. game dynamics: static game $\Rightarrow$ dynamic game

Extension I: Players

Multiple Receivers: Group Persuasion

- An increasing number of theoretical Foundation (Mathevet et al.2018; Arieli, Babichenko and Sandomirskiy,2021) justify the extension of concavification
- Alonso and Camara (2016b) propose a model with multiple receivers in the context of democracy and voting game
- Concavification can be extended naturally to this problem by considering procedures below:
  1. construct equilvalent single representative voter (the convex set of win set)
  2. implement BP to single representative voter
  3. ”segment” the signal to decompose the representative voter

Multiple Receivers: Group Persuasion

- policy implementation brings payoff δ^i_θ to voter i under state θ
- $a(q, \theta_i) = 1$ iff $\sum_{\theta \in \Theta} \delta^i_\theta q_\theta \geq 0$, ($q$ is the updated belief)
- at-least k consensus: $W_k = \{q \in \Delta(\Theta) | \sum_{i=1}^n a(q, \theta^i) \geq k\}$

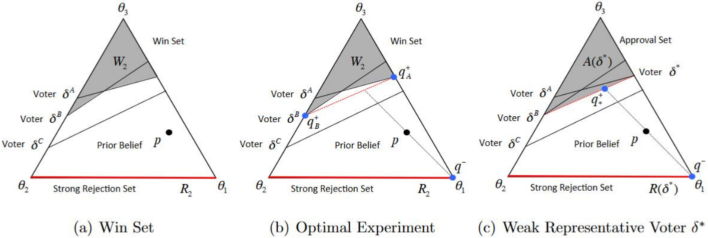


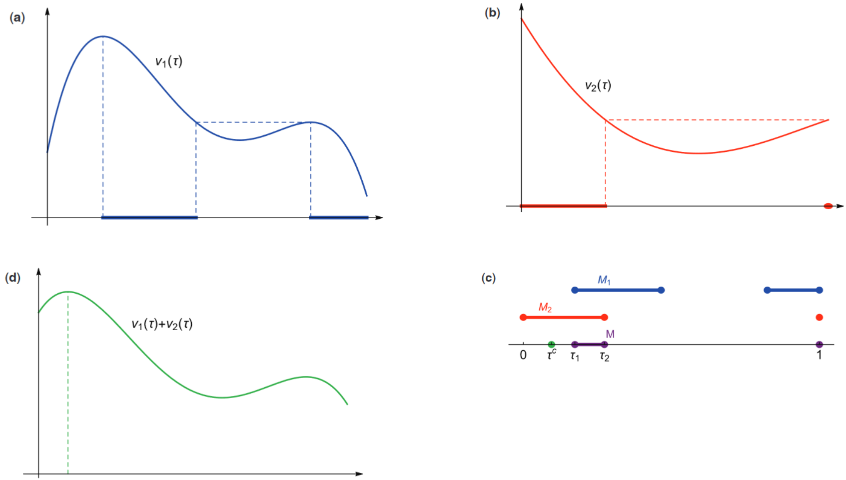
Figure 2: Optimal Experiment for Example 2

- Under a simple-majority rule, the politician's influence always makes a majority of voters weakly worse of

Multiple Senders: Competition Senders

- Kamenica (2017) shows collusive reveals less info than equil in Blackwell-Connected environment, where senders can pick any posterior belief of the receiver
- Introducing additional senders or decreasing the alignment of senders' preferences necessarily increases the amount of information revealed
- information struture τ is more informative than τ' (in Blackwell sense): $\tau \succeq \tau'$
- Nash equilibrium means no profitable deviation $\Rightarrow$ no profitable more informative $\tau \Rightarrow$ the intersection of such "sets"
- Nash equilibrium $\succeq$ collusive outcome

Multiple Senders: Competition Senders



Extension II: Action Space

Costly Information Design

- Several papers discuss the modelling of information production costs: processing costs (Lipnowski, Mathevet, and Wei, 2020), an axiomatic theory of information acquisition(Pomatto et.al,2023)
- Gentzkov and Kamenica (2014) considers a family of cost functions that is compatible with the concavification approach: proportional to expected reduction in entropy $c(\pi) = \mathbb{E}_{\langle \pi | \mu \rangle} [H(\mu) - H(\mu_s)]$ with constant scalar k

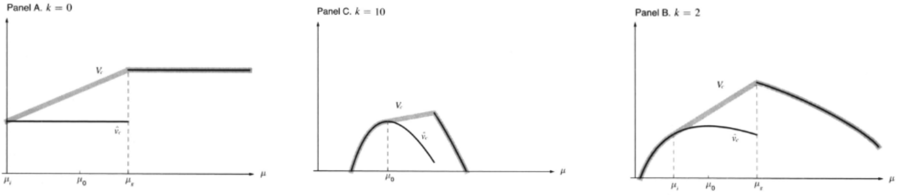


FIGURE 2. OPTIMAL INVESTIGATIONS

Extension III: Updating Rules

A General Framework of Non-bayesian Updating

- Clippel and Zhang(2022) considers a systematical non-bayesian updating with distortion function: $D_{\mu_0} : \Delta(\Omega) \to \Delta(\Omega)$ such that for all μ_0 $\mu^R(\cdot;\mu_0,\pi) = D_{\mu_0}(\mu^S(\cdot;\mu_0,\pi))$ for all signals π and all signal realizations s
  1. independent of the signal /neutrality
  2. independence of irrelevant signal realizations
- With the systematical distortion, the updating projection may not be convex, upsetting the revealation principle
- But concavification holds and we can derive some similar conclusions to Kamenica and Gentskov(2011)

Non-bayesian Updating: Examples

1. motivated updating (endogeneous choice): $D_{\mu_0}^{MU}(\nu) = \underset{\hat{\nu}\in\Gamma(\nu)}{\operatorname{argmax}} \mathcal{U}\left(\hat{\nu}, \nu, \nu^*\right)$
2. conservative updating: $D_{\mu_0}^{CB} = (1-\chi)\mu_0 + \chi\nu$

   affine updating: $D_{\mu_0}^{\chi,\nu^*} = (1-\chi)\nu^* + \chi\nu$
3. $\alpha - \beta$ updating: $\mu_s^R(\omega; \mu_0, \pi) = \frac{\pi(s|\omega)^\beta \mu_0(\omega)^\alpha}{\sum_{\omega'\in\Omega} \pi(s|\omega')^\beta \mu_0(\omega')^\alpha}$

   α: base rate negalect/overweighting prior, β: over/underinference

   $D_{\mu_0}^{\alpha,\beta}(\nu) = \frac{\nu^\beta \mu_0^{\alpha-\beta}}{\sum_{\omega'\in\Omega} \nu(\omega')^\beta \mu_0(\omega')^{\alpha-\beta}}$

RP: systematically distorted updating rules

- $\Omega = \{\omega_1, \omega_2, \omega_3\}$, μ_0. $A = \{a_1, a_2\}$
- $\mu_s^R(\omega; \mu_0, \pi) = \frac{\pi(s|\omega)^2 \mu_0(\omega)}{\sum_{\omega' \in \Omega} \pi(s|\omega')^2 \mu_0(\omega')}$, $D_{\mu_0}^{1,2}(\nu) = \frac{\nu^2 \mu_0^{-1}}{\sum_{\omega' \in \Omega} \nu(\omega')^2 \mu_0(\omega')^{-1}}$

	Signal			Receiver's Utility u	
	s_1	s_2	s_3	a_1	a_2
ω_1	1	0	0	ρ	0
ω_2	0	1	0	ρ	0
ω_3	φ	φ	$1 - 2\varphi$	0	1

- $0 < \phi < 0.5$ and $\phi^2 < \rho < 2\phi^2$, the receiver strictly prefers a_1 upon realizations s_1 and s_2 and strictly prefers a_2 upon s_3
- $S^{a_1} = \{s_1, s_2\}$ fails to recommend a_1

RP: systematically distorted updating rules

1. the optimal signal does require three realizations
2. the optimal signal involves only two realizations but cannot simply recommend an action that the receiver will follow
3. the revelation principle fails, but the optimal signal gives an incentive-compatible action recommendation

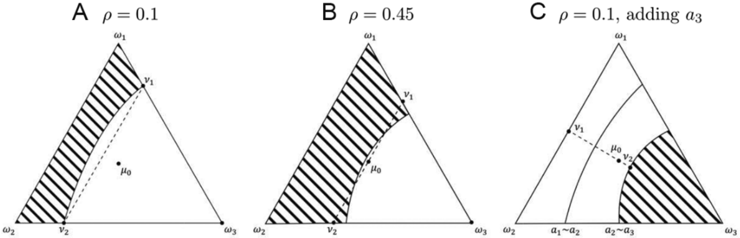


FIG. 1.—Illustration of sender's distorted indirect utility function $\hat{\nu}$. $\hat{\nu} = 1$ in the diagonally striped area (boundaries included); $\hat{\nu} = 0$ elsewhere.

Worldview Changing as Non-bayesian Updating

- Worldview: a proper subset of all states
  - "possible" states set $\mathcal{P}$ and "impossible" states set $\mathcal{I} = \Omega \backslash \mathcal{P}$
  - designer's prior: $\sigma \in \Delta(\Omega)$ with supp $\sigma = \Omega$
  - agent's prior: $\rho \in \Delta(\Omega)$ with supp $\rho = \mathcal{P} \subsetneq \Omega$
- Worldview Changing: only wheneye-opening, i.e. for every (s, π)
  1. resistance: $\pi(s \mid \omega) > 0$ for some $\omega \in \mathcal{P}$, (s, π) is "expected" for agent and "confirms"$\rho \Rightarrow$ updateρ using Bayes' rule
  2. worldview changing" $\pi(s \mid \omega) = 0$ for all $\omega \in \mathcal{P}$,(s, π) is "unexpected" for agent and "disproves" $\rho \Rightarrow$ change prior to ρ^1 with $supp \rho^1 = \Omega \Rightarrow$ update ρ^1 using Bayes' rule

Worldview Changing as Non-bayesian Updating

- specific properties: extremist; belief jump (discontinuity); concealment
- pooling as a strategy: pooling bad $\mathcal{P}$ with good $\mathcal{I}$ destroys good $\mathcal{I}$ and pooling good $\mathcal{P}$ with bad $\mathcal{I}$ benefits good $\mathcal{P}$

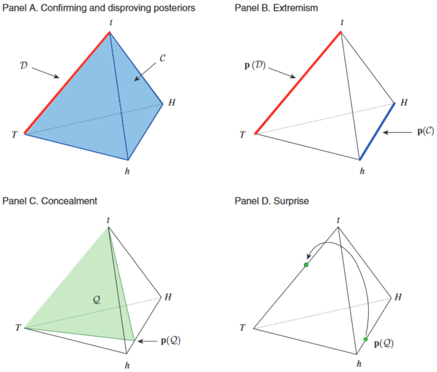


Figure 1. From Sender's to Receiver's Posterior

Worldview Changing as Non-bayesian Updating

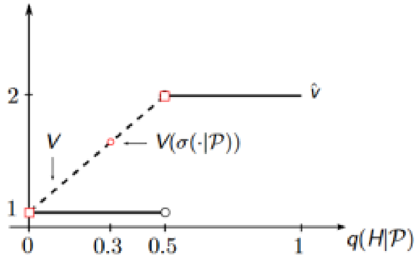


Figure: Confirming the agent

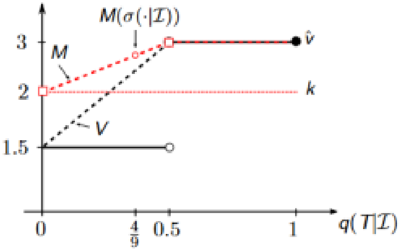


Figure: Disproving the agent

Extension IV: Game Rules

Partial Commitment Power

- Several papers analyse bayesian persuasion under partial commitment power (Min,2021) and test it with experiments (Fréchette et.al,2022)
- Lipnowski et.al(2022) proposes **a unified framework of cheap talk and bayesian persuasion** by defining an exogenenous variable $\chi \in [0, 1]$, measuring the commitment power of sender
- With χ, the sender reveals information as committed. with $1 - \chi$, the sender always reveals her most preferred information.

Partial Commitment Power

- Key insight: overcome the receiver‘s skepticism by releasing more information $\Rightarrow$ beneficial to receiver

  small decrease in credibilty $\Rightarrow$ sharp decrease in sender‘s payoff under some situations

- Considering an investment game between government and firms

ω	large	small	no
g	2	1	0
b	-1	-1/4	0

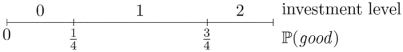


FIG. 1.—Firm’s best response in central bank example.

- $\chi = 1$: $\xi_1^*(g \mid good) = 3/4, \xi_1^*(b \mid bad) = 3/4$
- $\chi = 2/3$: $\xi_1^*(g \mid good) = 1, \xi_1^*(b \mid bad) = 1$

Equilirbium Selection: Adversial Information Design

- There exists many outstanding papers considering the equil prediction/revenue guarantee across all info structures (Bergeman et.al,2017; Du,2018) or max-min profits across equil and all info structures (Brooks and Du,2018)
- Dworczak and Pavan(2022) considers the adversial information design in two meanings: **the worst equil selection by receiver** and **worst signal provision by nature/victim**
  1. the worst equil selection by receiver: revise the equil action set of receivers
  2. worst signal provision by nature/victim: a reverse concavification
- Key Conclusion: the convex combination of payoffs of full revealation signals should be the lower bound of lower convex closure (Cautious with concavification!)
- Advances:a more general equil selection (Lipnowski, Ravid and Shishkin,2022)

Equilirbium Selection: Adversial Information Design

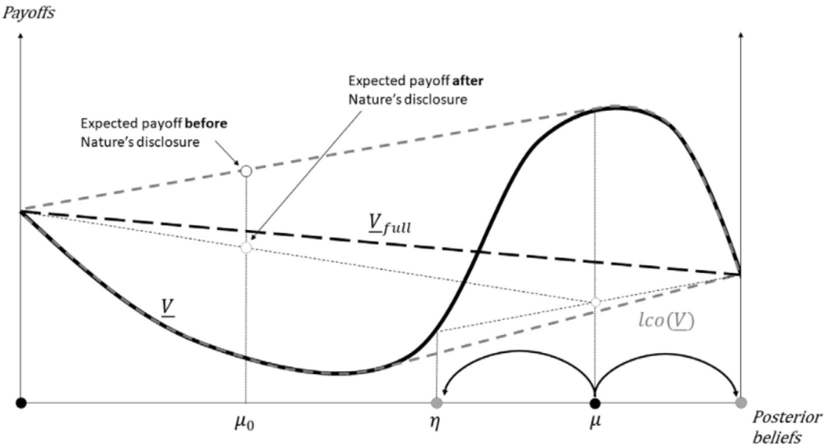


FIGURE 1.—Illustration of Proposition 1.

Extension V: Game Dynamics

Dynamic Information Design

- Insight: Information can be used as **an incentive device to reward behavior** over time.
- Moving the Goalshots: Ely and Szydlowski (2017) investigates the optimal dynamic design of information about difficulty of a task to incentivize sustained effort
- Suspense and Surprise: Ely (2015) investigates the optimal dynamic design of two non-instrumental information: suspense and surprise, i.e. (expected) variance along the belief path
- Foundation: Doval and Ely (2020) investigates the optimal dynamic design of information about evolving state when agents choose action every period

Applications

1. Introduction
2. Benchmark Model
   - Setup
   - Belief Approach
   - Other Methods
3. Extensions
   - Extension I: Players
   - Extension II: Action Space
   - Extension III: Updating Rules
   - Extension IV: Game Rules
   - Extension V: Game Dynamics
4. **Applications**
5. What I wish I had time to cover...

Applications

Examples of such problems (Kamenica,2019):

- financial sector stress tests (Goldstein and Leitner,2018; Inostroza-Pavan,2018; Orlov et al.,2018)
- grading in schools (Boleslavsky-Cotton,2015; Ostrovsky-Schwarz,2010)
- employee feedback (Habibi,2018; Smolin 2017)
- law enforcement deployment (Hernandez-Neeman 2017; Lazear,2006; Rabinovich et al.,2015)
- censorship (Gehlbach-Sonin,2014)
- **entertainment (Ely et al.,2015)**
- **voter coalition formation (Alonso-Camara 2016)**
- research procurement (Yoder 2018)
- medical research or testing (Kolotilin 2015,Schweizer-Szech 2019)
- matching platforms (Romanyuk-Smolin 2019)
- **price discrimination (Bergemann et al.,2015)**
- insurance (Garcia-Tsur,2018)
- transparency in organizations (Jehiel,2015)
- contest design (Zhang and Zhou,2016)

What I wish I had time to cover...

What I wish I had time to cover...

- What I wish I had time to cover:
  1. a general framework/equivalence to other topics(Kolotilin and Zapechelnyuk (2018); Kleiner, Moldovanu, and Strack (2020))
  2. the sources of (lack of) committment power (Nguyan and Tan,2021)
  3. incoporation of Contract Design/Auction/Network (Boleslavsky and Kim,2018)
  4. simplification techniques (Xiaoyu Cheng,2021; Lipnowski and Mathevet,2017)
  5. bayesian persuasion with mediators (Arieli, Babichenko and Sandomirskiy,2023)
  6. some empirical/experimental studies
  7. many oustanding researches I omitted......