Robust scoring rules

Elias Tsakas

Maastricht University

June 2019
BGSE Summer Forum
Bounded Rationality, Cognition and Strategic Uncertainty

Roadmap

1. Motivation and Contribution
2. Formal model
3. First result: Exact robustness
4. Extensions: Approximate robustness

Roadmap

1. Motivation and Contribution
2. Formal model
3. First result: Exact robustness
4. Extensions: Approximate robustness

Background

General problem: How to elicit latent subjective beliefs?

- **Usual answer:** Proper scoring rules.
- **Methodological problem (Heisenberg):** Monetary incentives (provided by the scoring rule) may affect the very same beliefs we want to elicit.

Background

General problem: How to elicit latent subjective beliefs?

- **Usual answer:** Proper scoring rules.
- **Methodological problem (Heisenberg):** Monetary incentives (provided by the scoring rule) may affect the very same beliefs we want to elicit.

Background

General problem: How to elicit latent subjective beliefs?

- **Usual answer:** Proper scoring rules.
- **Methodological problem (Heisenberg):** Monetary incentives (provided by the scoring rule) may affect the very same beliefs we want to elicit.

Example: Eliciting population beliefs

- We are interested in distribution of beliefs in a population, e.g.,
  - Political campaign (population of voters)
  - Marketing campaign (population of consumers)
- Importantly, we are not interested in the actual state.
- Three steps to estimate the population beliefs:
  1. draw a representative sample from the population,
  2. elicit individual beliefs from each subject in the sample,
  3. use frequency of elicited beliefs as an estimate for population.
- **Problem:** If subjects in the sample respond to the incentives we provide them, by acquiring information, then we obtain biased estimate of population beliefs.

Example: Eliciting population beliefs

- We are interested in distribution of beliefs in a population, e.g.,
  - Political campaign (population of voters)
  - Marketing campaign (population of consumers)
- Importantly, we are not interested in the actual state.
- Three steps to estimate the population beliefs:
  1. draw a representative sample from the population,
  2. elicit individual beliefs from each subject in the sample,
  3. use frequency of elicited beliefs as an estimate for population.
- **Problem:** If subjects in the sample respond to the incentives we provide them, by acquiring information, then we obtain biased estimate of population beliefs.

Example: Eliciting population beliefs

- We are interested in distribution of beliefs in a population, e.g.,
  - Political campaign (population of voters)
  - Marketing campaign (population of consumers)
- Importantly, we are not interested in the actual state.
- Three steps to estimate the population beliefs:
  1. draw a representative sample from the population,
  2. elicit individual beliefs from each subject in the sample,
  3. use frequency of elicited beliefs as an estimate for population.
- **Problem:** If subjects in the sample respond to the incentives we provide them, by acquiring information, then we obtain biased estimate of population beliefs.

Example: Eliciting population beliefs

- We are interested in distribution of beliefs in a population, e.g.,
  - Political campaign (population of voters)
  - Marketing campaign (population of consumers)
- Importantly, we are not interested in the actual state.
- Three steps to estimate the population beliefs:
  1. draw a representative sample from the population,
  2. elicit individual beliefs from each subject in the sample,
  3. use frequency of elicited beliefs as an estimate for population.
- **Problem:** If subjects in the sample respond to the incentives we provide them, by acquiring information, then we obtain biased estimate of population beliefs.

Example: Eliciting population beliefs

- We are interested in distribution of beliefs in a population, e.g.,
  - Political campaign (population of voters)
  - Marketing campaign (population of consumers)
- Importantly, we are not interested in the actual state.
- Three steps to estimate the population beliefs:
  1. draw a representative sample from the population,
  2. elicit individual beliefs from each subject in the sample,
  3. use frequency of elicited beliefs as an estimate for population.
- **Problem:** If subjects in the sample respond to the incentives we provide them, by acquiring information, then we obtain biased estimate of population beliefs.

Example: Eliciting population beliefs

- We are interested in distribution of beliefs in a population, e.g.,
  - Political campaign (population of voters)
  - Marketing campaign (population of consumers)
- Importantly, we are not interested in the actual state.
- Three steps to estimate the population beliefs:
  1. draw a representative sample from the population,
  2. elicit individual beliefs from each subject in the sample,
  3. use frequency of elicited beliefs as an estimate for population.
- **Problem:** If subjects in the sample respond to the incentives we provide them, by acquiring information, then we obtain biased estimate of population beliefs.

Research question and preview of results

Can we elicit the beliefs that the subject would have had, if the elicitation task had not taken place?

- These are called prior beliefs.
- Yes, under standard mild assumptions.
- If we accept small mistakes, any proper scoring rules would work.

Research question and preview of results

Can we elicit the beliefs that the subject would have had, if the elicitation task had not taken place?

- These are called prior beliefs.
- Yes, under standard mild assumptions.
- If we accept small mistakes, any proper scoring rules would work.

Research question and preview of results

Can we elicit the beliefs that the subject would have had, if the elicitation task had not taken place?

- These are called prior beliefs.
- Yes, under standard mild assumptions.
- If we accept small mistakes, any proper scoring rules would work.

Literature(s)

INTERSECTION OF TWO LITERATURES:

- Incentivized belief elicitation (scoring rules)
- Rational inattention / Costly information acquisition

Roadmap

Preliminaries

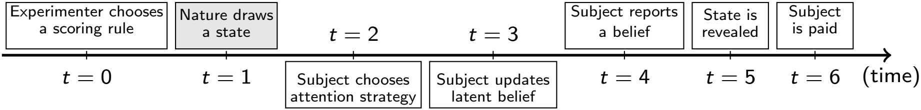
FUNDAMENTALS:

- Binary state space: $\Omega = \{\omega_0, \omega_1\}$
- Latent subjective belief (of ω_0 occurring): $\mu \in [0, 1]$
- (Non-verifiable) self-report: $r \in [0, 1]$

Elicitation Mechanisms:

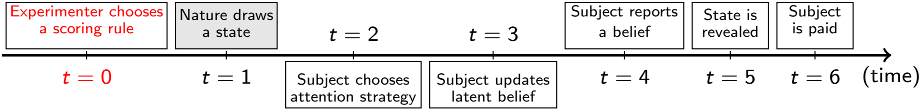
- Scoring rule: $\boxed{S : [0, 1] \times \Omega \to \mathbb{R}}$
- Payment depends on self-report and state realization.
- Payment is in monetary payoffs.

Purpose and timeline of our mechanism



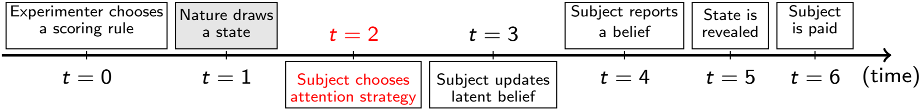
- The experimenter wants to elicit prior beliefs.
- So, she wants to design a scoring rule, such that the subject
  1. does not acquire any information,
  2. so that he does not update his beliefs, and
  3. subsequently he reports truthfully.
- We will call this, a robust scoring rule.

Purpose and timeline of our mechanism



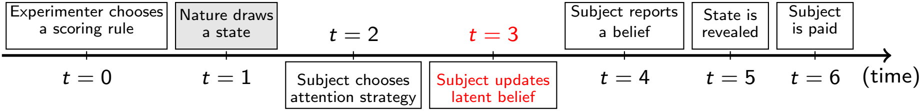
- The experimenter wants to elicit prior beliefs.
- So, she wants to design a scoring rule, such that the subject
  1. does not acquire any information,
  2. so that he does not update his beliefs, and
  3. subsequently he reports truthfully.
- We will call this, a robust scoring rule.

Purpose and timeline of our mechanism



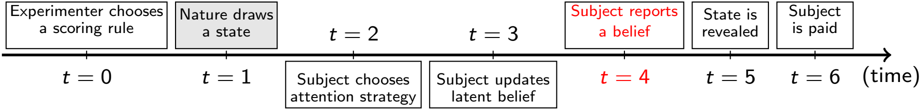
- The experimenter wants to elicit prior beliefs.
- So, she wants to design a scoring rule, such that the subject
  1. does not acquire any information,
  2. so that he does not update his beliefs, and
  3. subsequently he reports truthfully.
- We will call this, a robust scoring rule.

Purpose and timeline of our mechanism



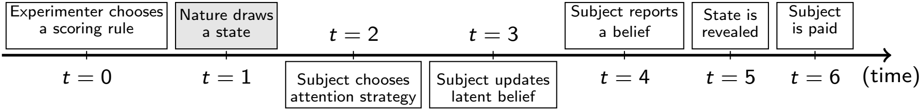
- The experimenter wants to elicit prior beliefs.
- So, she wants to design a scoring rule, such that the subject
  1. does not acquire any information,
  2. so that he does not update his beliefs, and
  3. subsequently he reports truthfully.
- We will call this, a robust scoring rule.

Purpose and timeline of our mechanism



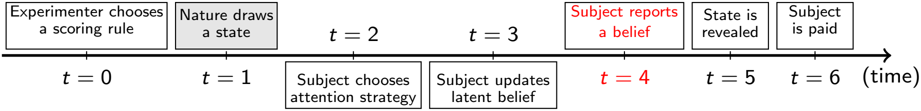
- The experimenter wants to elicit prior beliefs.
- So, she wants to design a scoring rule, such that the subject
  1. does not acquire any information,
  2. so that he does not update his beliefs, and
  3. subsequently he reports truthfully.
- We will call this, a robust scoring rule.

Purpose and timeline of our mechanism



- The experimenter wants to elicit prior beliefs.
- So, she wants to design a scoring rule, such that the subject
  1. does not acquire any information,
  2. so that he does not update his beliefs, and
  3. subsequently he reports truthfully.
- We will call this, a robust scoring rule.

Proper scoring rules



- Begin with the second step (i.e., "report truthfully").

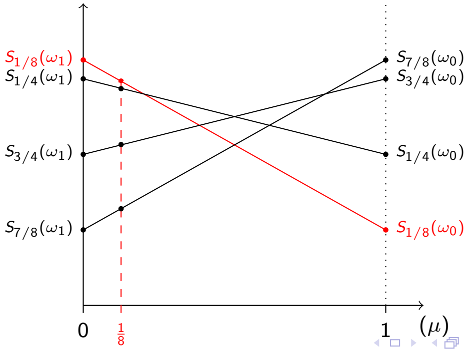
Proper scoring rules

- **Definition:** Strictly dominant to report truthfully for all beliefs

$$\mathbb{E}_\mu(S_\mu) > \mathbb{E}_\mu(S_r) \text{ for all } r \neq \mu \text{ and all } \mu \in [0,1]$$

- **Characterization** (Savage, 1971): Define $\phi(\mu) := \mathbb{E}_\mu(S_\mu)$

S proper $\Leftrightarrow$ ϕ strictly convex and subdifferentiable.



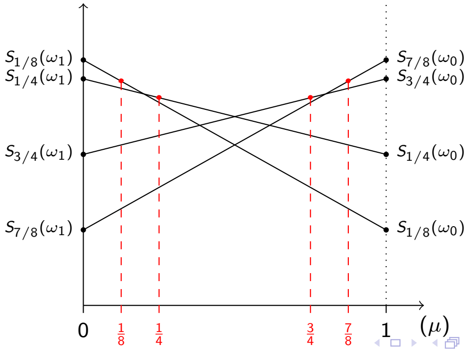
Proper scoring rules

- **Definition:** Strictly dominant to report truthfully for all beliefs

$$\mathbb{E}_\mu(S_\mu) > \mathbb{E}_\mu(S_r) \text{ for all } r \neq \mu \text{ and all } \mu \in [0,1]$$

- **Characterization** (Savage, 1971): Define $\phi(\mu) := \mathbb{E}_\mu(S_\mu)$

$$S \text{ proper} \Leftrightarrow \phi \text{ strictly convex and subdifferentiable.}$$



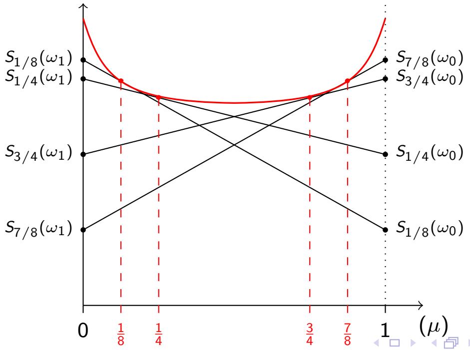
Proper scoring rules

- **Definition:** Strictly dominant to report truthfully for all beliefs

$$\boxed{\mathbb{E}_\mu(S_\mu) > \mathbb{E}_\mu(S_r) \text{ for all } r \neq \mu \text{ and all } \mu \in [0,1]}$$

- **Characterization** (Savage, 1971): Define $\phi(\mu) := \mathbb{E}_\mu(S_\mu)$

$$\boxed{S \text{ proper} \Leftrightarrow \phi \text{ strictly convex and subdifferentiable.}}$$



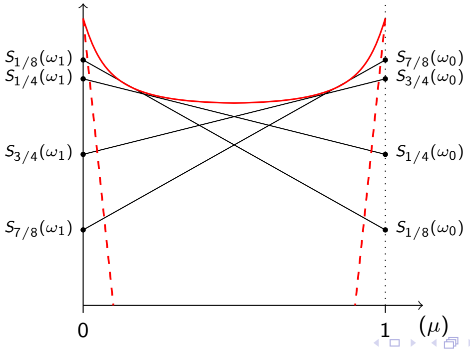
Proper scoring rules

- **Definition:** Strictly dominant to report truthfully for all beliefs

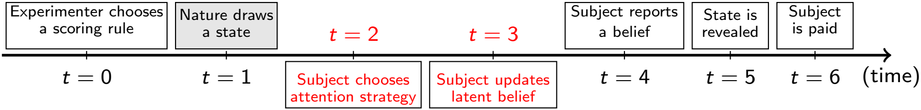
$$\boxed{\mathbb{E}_\mu(S_\mu) > \mathbb{E}_\mu(S_r) \text{ for all } r \neq \mu \text{ and all } \mu \in [0,1]}$$

- **Characterization** (Savage, 1971): Define $\phi(\mu) := \mathbb{E}_\mu(S_\mu)$

$$\boxed{S \text{ proper} \Leftrightarrow \phi \text{ strictly convex and subdifferentiable.}}$$



Attention



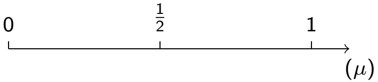
- Continue with the first step (i.e., "not acquiring any information").

Attention strategies

- An **attention strategy** is modelled with a Bayesian signal, $\sigma : \Omega \to \Delta(S)$ chosen by the subject.
- Given a prior $\mu \in [0,1]$, each feasible attention strategy is characterized by a (mean-preserving) distribution of posteriors:

$$\boxed{\pi \in \Delta([0,1]) \text{ such that } \mu = \mathbb{E}_\pi(\nu)}.$$

- The set of feasible attention strategies is denoted by $\Pi(\mu)$.
- Important special cases:
  - No-attention strategy: $\hat{\mu} \in \Pi(\mu)$ puts probability 1 to μ.
  - Perfectly informative strategy: $\pi^*_\mu \in \Pi(\mu)$ puts probability 1 to $\{0,1\}$.
- Attention has benefits and costs.

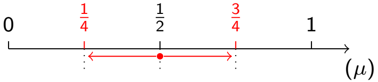


Attention strategies

- An **attention strategy** is modelled with a Bayesian signal, $\sigma : \Omega \to \Delta(S)$ chosen by the subject.
- Given a prior $\mu \in [0,1]$, each feasible attention strategy is characterized by a (mean-preserving) distribution of posteriors:

$$\boxed{\pi \in \Delta([0,1]) \text{ such that } \mu = \mathbb{E}_\pi(\nu)}.$$

- The set of feasible attention strategies is denoted by $\Pi(\mu)$.
- Important special cases:
  - No-attention strategy: $\hat{\mu} \in \Pi(\mu)$ puts probability 1 to μ.
  - Perfectly informative strategy: $\pi^*_\mu \in \Pi(\mu)$ puts probability 1 to $\{0,1\}$.
- Attention has benefits and costs.

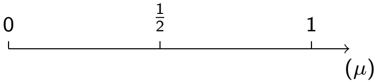


Attention strategies

- An **attention strategy** is modelled with a Bayesian signal, $\sigma : \Omega \to \Delta(S)$ chosen by the subject.
- Given a prior $\mu \in [0, 1]$, each feasible attention strategy is characterized by a (mean-preserving) distribution of posteriors:

$$\boxed{\pi \in \Delta([0,1]) \text{ such that } \mu = \mathbb{E}_\pi(\nu)}.$$

- The set of feasible attention strategies is denoted by $\Pi(\mu)$.
- Important special cases:
    - No-attention strategy: $\hat{\mu} \in \Pi(\mu)$ puts probability 1 to μ.
    - Perfectly informative strategy: $\pi^*_\mu \in \Pi(\mu)$ puts probability 1 to $\{0, 1\}$.
- Attention has benefits and costs.

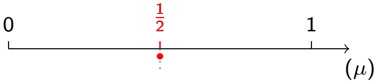


Attention strategies

- An **attention strategy** is modelled with a Bayesian signal, $\sigma : \Omega \to \Delta(S)$ chosen by the subject.
- Given a prior $\mu \in [0,1]$, each feasible attention strategy is characterized by a (mean-preserving) distribution of posteriors:

$$\boxed{\pi \in \Delta([0,1]) \text{ such that } \mu = \mathbb{E}_\pi(\nu)}.$$

- The set of feasible attention strategies is denoted by $\Pi(\mu)$.
- Important special cases:
  - No-attention strategy: $\hat{\mu} \in \Pi(\mu)$ puts probability 1 to μ.
  - Perfectly informative strategy: $\pi^*_\mu \in \Pi(\mu)$ puts probability 1 to $\{0,1\}$.
- Attention has benefits and costs.

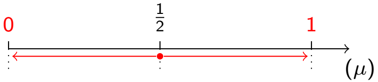


Attention strategies

- An **attention strategy** is modelled with a Bayesian signal, $\sigma : \Omega \to \Delta(S)$ chosen by the subject.
- Given a prior $\mu \in [0,1]$, each feasible attention strategy is characterized by a (mean-preserving) distribution of posteriors:

$$\boxed{\pi \in \Delta([0,1]) \text{ such that } \mu = \mathbb{E}_\pi(\nu)}.$$

- The set of feasible attention strategies is denoted by $\Pi(\mu)$.
- Important special cases:
  - No-attention strategy: $\hat{\mu} \in \Pi(\mu)$ puts probability 1 to μ.
  - Perfectly informative strategy: $\pi^*_\mu \in \Pi(\mu)$ puts probability 1 to $\{0,1\}$.
- Attention has benefits and costs.

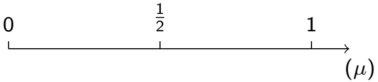


Attention strategies

- An **attention strategy** is modelled with a Bayesian signal, $\sigma : \Omega \to \Delta(S)$ chosen by the subject.
- Given a prior $\mu \in [0,1]$, each feasible attention strategy is characterized by a (mean-preserving) distribution of posteriors:

$$\boxed{\pi \in \Delta([0,1]) \text{ such that } \mu = \mathbb{E}_\pi(\nu)}.$$

- The set of feasible attention strategies is denoted by $\Pi(\mu)$.
- Important special cases:
  - No-attention strategy: $\hat{\mu} \in \Pi(\mu)$ puts probability 1 to μ.
  - Perfectly informative strategy: $\pi^*_\mu \in \Pi(\mu)$ puts probability 1 to $\{0,1\}$.
- Attention has benefits and costs.

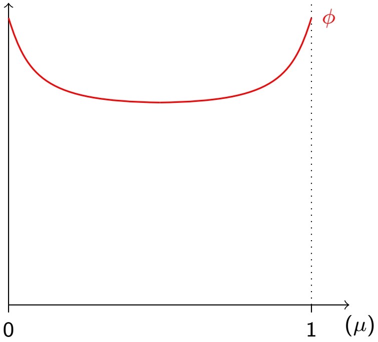


Benefit of attention

- Fix a proper scoring rule ϕ.
- For prior μ and attention $\pi \in \Pi(\mu)$, the expected benefit is

$$\boxed{B_\phi(\pi) = \langle \phi, \pi \rangle - \phi(\mu)}$$

- Every attention strategy yields a strictly positive expected benefit
- The more convex ϕ is, the stronger the incentives.

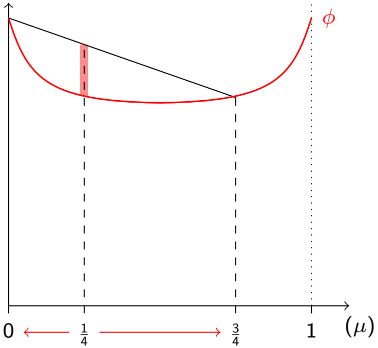


Benefit of attention

- Fix a proper scoring rule ϕ.
- For prior μ and attention $\pi \in \Pi(\mu)$, the expected benefit is

$$\boxed{B_\phi(\pi) = \langle \phi, \pi \rangle - \phi(\mu)}$$

- Every attention strategy yields a strictly positive expected benefit
- The more convex ϕ is, the stronger the incentives.

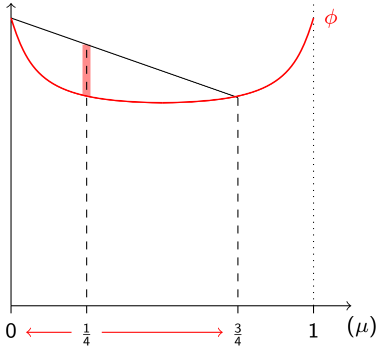


Benefit of attention

- Fix a proper scoring rule ϕ.
- For prior μ and attention $\pi \in \Pi(\mu)$, the expected benefit is

$$\boxed{B_\phi(\pi) = \langle \phi, \pi \rangle - \phi(\mu)}$$

- Every attention strategy yields a strictly positive expected benefit
- The more convex ϕ is, the stronger the incentives.

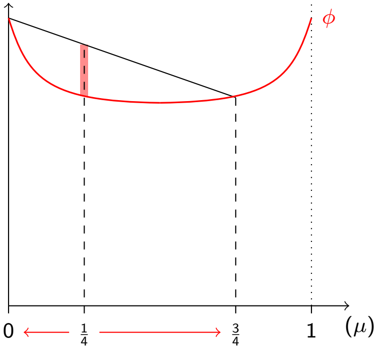


Benefit of attention

- Fix a proper scoring rule ϕ.
- For prior μ and attention $\pi \in \Pi(\mu)$, the expected benefit is

$$\boxed{B_\phi(\pi) = \langle \phi, \pi \rangle - \phi(\mu)}$$

- Every attention strategy yields a strictly positive expected benefit
- The more convex ϕ is, the stronger the incentives.



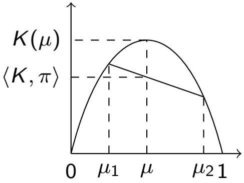
Cost of attention

$$C : \Delta([0,1]) \to \mathbb{R}_+$$

- Costs are assumed to satisfy **posterior separability**:

There is a strictly concave function $K : [0,1] \to \mathbb{R}$ such that
$$C(\pi) = K(\mu) - \langle K, \pi \rangle$$
for every $\pi \in \Pi(\mu)$ and every $\mu \in [0,1]$.

- Supporting experimental evidence (Dean & Neligh, 2017)
- Solid theoretical foundations (see Appendix)
- Usual applications (entropic costs)



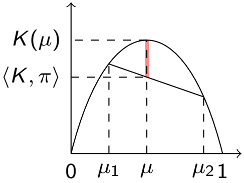
Cost of attention

$$\boxed{C : \Delta([0,1]) \to \mathbb{R}_+}$$

- Costs are assumed to satisfy **posterior separability**:

> There is a strictly concave function $K : [0,1] \to \mathbb{R}$ such that
> $$C(\pi) = K(\mu) - \langle K, \pi \rangle$$
> for every $\pi \in \Pi(\mu)$ and every $\mu \in [0,1]$.

- Supporting experimental evidence (Dean & Neligh, 2017)
- Solid theoretical foundations (see Appendix)
- Usual applications (entropic costs)



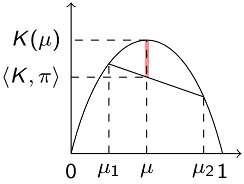
Cost of attention

$$\boxed{C : \Delta([0,1]) \to \mathbb{R}_+}$$

- Costs are assumed to satisfy **posterior separability**:

> There is a strictly concave function $K : [0,1] \to \mathbb{R}$ such that
> $$C(\pi) = K(\mu) - \langle K, \pi \rangle$$
> for every $\pi \in \Pi(\mu)$ and every $\mu \in [0,1]$.

- Supporting experimental evidence (Dean & Neligh, 2017)
- Solid theoretical foundations (see Appendix)
- Usual applications (entropic costs)



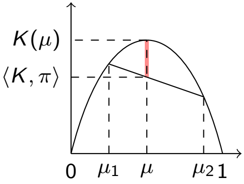
Cost of attention

$$\boxed{C : \Delta([0,1]) \to \mathbb{R}_+}$$

- Costs are assumed to satisfy **posterior separability**:

There is a strictly concave function $K : [0,1] \to \mathbb{R}$ such that
$$C(\pi) = K(\mu) - \langle K, \pi \rangle$$
for every $\pi \in \Pi(\mu)$ and every $\mu \in [0,1]$.

- Supporting experimental evidence (Dean & Neligh, 2017)
- Solid theoretical foundations (see Appendix)
- Usual applications (entropic costs)



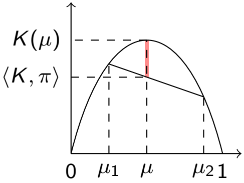
Cost of attention

$$\boxed{C : \Delta([0,1]) \to \mathbb{R}_+}$$

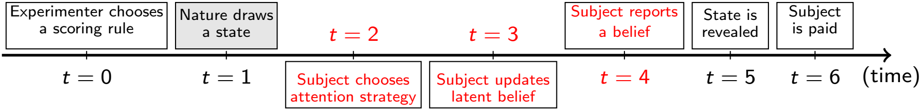
- Costs are assumed to satisfy **posterior separability**:

There is a strictly concave function $K : [0,1] \to \mathbb{R}$ such that
$$C(\pi) = K(\mu) - \langle K, \pi \rangle$$
for every $\pi \in \Pi(\mu)$ and every $\mu \in [0,1]$.

- Supporting experimental evidence (Dean & Neligh, 2017)
- Solid theoretical foundations (see Appendix)
- Usual applications (entropic costs)



Eliciting prior belief



- Put the two steps together (i.e., “not acquire any information” and “report truthfully”).

Robust scoring rule

- Value of attention given a scoring rule:

$$V_\phi(\pi) := B_\phi(\pi) - C(\pi)$$

- **Robust scoring rule:** For every prior $\mu \in [0,1]$, it is simultaneously strictly dominant:
  1. not to acquire any information, and
  2. to report truthfully.

$$V_\phi(\hat{\mu}) > V_\phi(\pi) \text{ for all } \pi \in \hat{\Pi}(\mu), \text{ with } \phi \text{ proper}$$

- Intuitively, the incentives should be strong enough to tell the truth, but not too strong so that the expected benefit from acquiring information offsets the cost.
- That is, "ϕ must be strictly convex, but not too convex".

Robust scoring rule

- Value of attention given a scoring rule:

$$\boxed{V_\phi(\pi) := B_\phi(\pi) - C(\pi)}$$

- **Robust scoring rule:** For every prior $\mu \in [0,1]$, it is simultaneously strictly dominant:
  1. not to acquire any information, and
  2. to report truthfully.

$$\boxed{V_\phi(\hat{\mu}) > V_\phi(\pi) \text{ for all } \pi \in \hat{\Pi}(\mu), \text{ with } \phi \text{ proper}}$$

- Intuitively, the incentives should be strong enough to tell the truth, but not too strong so that the expected benefit from acquiring information offsets the cost.
- That is, "ϕ must be strictly convex, but not too convex".

Robust scoring rule

- Value of attention given a scoring rule:

$$\boxed{V_\phi(\pi) := B_\phi(\pi) - C(\pi)}$$

- **Robust scoring rule:** For every prior $\mu \in [0,1]$, it is simultaneously strictly dominant:
  1. not to acquire any information, and
  2. to report truthfully.

$$\boxed{V_\phi(\hat{\mu}) > V_\phi(\pi) \text{ for all } \pi \in \hat{\Pi}(\mu), \text{ with } \phi \text{ proper}}$$

- Intuitively, the incentives should be strong enough to tell the truth, but not too strong so that the expected benefit from acquiring information offsets the cost.
- That is, "ϕ must be strictly convex, but not too convex".

Robust scoring rule

- Value of attention given a scoring rule:

$$\boxed{V_\phi(\pi) := B_\phi(\pi) - C(\pi)}$$

- **Robust scoring rule:** For every prior $\mu \in [0,1]$, it is simultaneously strictly dominant:
  1. not to acquire any information, and
  2. to report truthfully.

$$\boxed{V_\phi(\hat{\mu}) > V_\phi(\pi) \text{ for all } \pi \in \hat{\Pi}(\mu), \text{ with } \phi \text{ proper}}$$

- Intuitively, the incentives should be strong enough to tell the truth, but not too strong so that the expected benefit from acquiring information offsets the cost.
- That is, "ϕ must be strictly convex, but not too convex".

Robust scoring rule

- Value of attention given a scoring rule:

$$\boxed{V_\phi(\pi) := B_\phi(\pi) - C(\pi)}$$

- **Robust scoring rule:** For every prior $\mu \in [0,1]$, it is simultaneously strictly dominant:
  1. not to acquire any information, and
  2. to report truthfully.

$$\boxed{V_\phi(\hat{\mu}) > V_\phi(\pi) \text{ for all } \pi \in \hat{\Pi}(\mu), \text{ with } \phi \text{ proper}}$$

- Intuitively, the incentives should be strong enough to tell the truth, but not too strong so that the expected benefit from acquiring information offsets the cost.
- That is, "ϕ must be strictly convex, but not too convex".

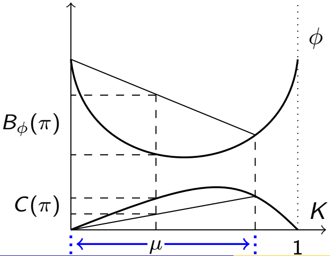
Optimal attention strategy: Concavification

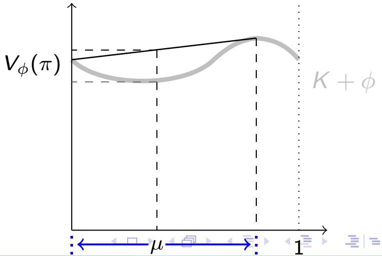
- Value of $\pi \in \Pi(\mu)$ (for a proper ϕ and posterior-separable K):

$$\begin{aligned} V_\phi(\pi) &= B_\phi(\pi) - C(\pi) \\ &= \big(\langle \phi, \pi \rangle - \phi(\mu)\big) - \big(K(\mu) - \langle K, \pi \rangle\big) \\ &= \langle K + \phi, \pi \rangle + (K + \phi) \end{aligned}$$

- Optimal π given by concave closure of $K + \phi$ (Aumann & Maschler, 1995).

$\boxed{\phi \text{ robust} \Leftrightarrow K + \phi \text{ strictly concave}}$





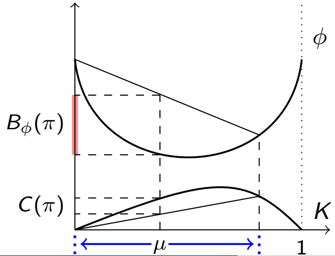
Optimal attention strategy: Concavification

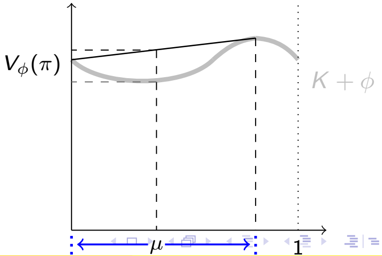
- Value of $\pi \in \Pi(\mu)$ (for a proper ϕ and posterior-separable K):

$$\begin{aligned} V_\phi(\pi) &= B_\phi(\pi) - C(\pi) \\ &= \big(\langle \phi, \pi \rangle - \phi(\mu)\big) - \big(K(\mu) - \langle K, \pi \rangle\big) \\ &= \langle K + \phi, \pi \rangle + (K + \phi) \end{aligned}$$

- Optimal π given by concave closure of $K + \phi$ (Aumann & Maschler, 1995).

ϕ robust $\Leftrightarrow$ $K + \phi$ strictly concave





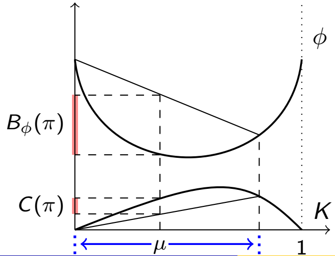
Optimal attention strategy: Concavification

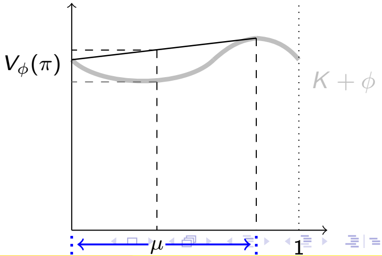
- Value of $\pi \in \Pi(\mu)$ (for a proper ϕ and posterior-separable K):

$$\begin{aligned} V_\phi(\pi) &= B_\phi(\pi) - C(\pi) \\ &= \big(\langle \phi, \pi\rangle - \phi(\mu)\big) - \big(K(\mu) - \langle K, \pi\rangle\big) \\ &= \langle K + \phi, \pi\rangle + (K + \phi) \end{aligned}$$

- Optimal π given by concave closure of $K + \phi$ (Aumann & Maschler, 1995).

ϕ robust $\Leftrightarrow$ $K + \phi$ strictly concave





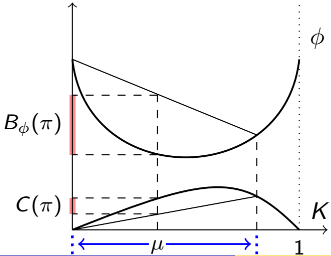
Optimal attention strategy: Concavification

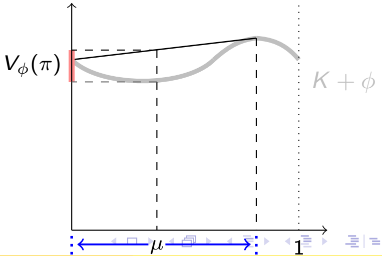
- Value of $\pi \in \Pi(\mu)$ (for a proper ϕ and posterior-separable K):

$$\begin{aligned} V_\phi(\pi) &= B_\phi(\pi) - C(\pi) \\ &= \big(\langle \phi, \pi \rangle - \phi(\mu)\big) - \big(K(\mu) - \langle K, \pi \rangle\big) \\ &= \langle K + \phi, \pi \rangle + (K + \phi) \end{aligned}$$

- Optimal π given by concave closure of $K + \phi$ (Aumann & Maschler, 1995).

ϕ robust $\Leftrightarrow$ $K + \phi$ strictly concave





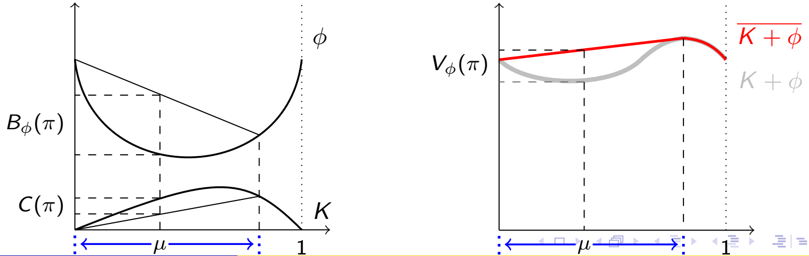
Optimal attention strategy: Concavification

- Value of $\pi \in \Pi(\mu)$ (for a proper ϕ and posterior-separable K):

$$\begin{aligned} V_\phi(\pi) &= B_\phi(\pi) - C(\pi) \\ &= \big(\langle \phi, \pi \rangle - \phi(\mu)\big) - \big(K(\mu) - \langle K, \pi \rangle\big) \\ &= \langle K + \phi, \pi \rangle + (K + \phi) \end{aligned}$$

- Optimal π given by concave closure of $K + \phi$ (Aumann & Maschler, 1995).

ϕ robust $\Leftrightarrow$ $K + \phi$ strictly concave



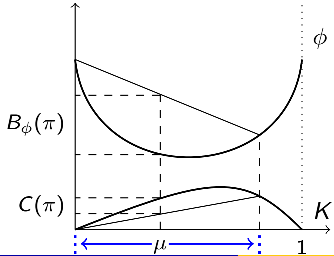
Optimal attention strategy: Concavification

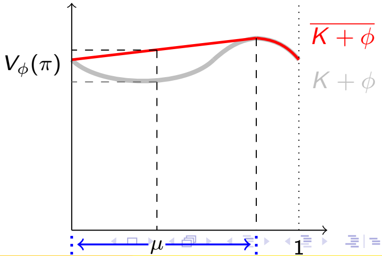
- Value of $\pi \in \Pi(\mu)$ (for a proper ϕ and posterior-separable K):

$$\begin{aligned} V_\phi(\pi) &= B_\phi(\pi) - C(\pi) \\ &= \big(\langle \phi, \pi\rangle - \phi(\mu)\big) - \big(K(\mu) - \langle K, \pi\rangle\big) \\ &= \langle K + \phi, \pi\rangle + (K + \phi) \end{aligned}$$

- Optimal π given by concave closure of $K + \phi$ (Aumann & Maschler, 1995).

ϕ robust $\Leftrightarrow$ $K + \phi$ strictly concave





Example: Robust QSR under entropic costs

- **Quadratic scoring rule:** $\phi(\mu) = \alpha - \beta\mu(1-\mu)$
- **Entropic costs:** $K(\mu) = -\kappa\big(\mu \log \mu + (1-\mu)\log(1-\mu)\big)$

$$\begin{aligned} \phi \text{ robust} \quad &\Leftrightarrow \quad K + \phi \text{ strictly concave} \\ &\Leftrightarrow \quad \beta \leq 2\kappa. \end{aligned}$$

Roadmap

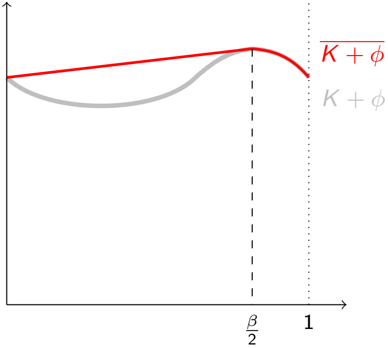
Existence

Theorem

If the cost function is posterior-separable, there exists a robust scoring rule.

Non-existence of robust QSR

- **Quadratic scoring rule:** $\phi(\mu) = \alpha - \beta\mu(1-\mu)$
- **Costs:** $K(\mu) = \mu - \mu^3$
- For all $\beta > 0$ and all $\mu \in (0, \beta/2)$, the subject updates his beliefs.



Roadmap

Additional questions

- What if the cost function is not known to the experimenter?
- What can we achieve with well-known scoring rules (e.g., quadratic scoring rule)?
- We answer both at once, by considering approximate robustness.

Additional questions

- What if the cost function is not known to the experimenter?
- What can we achieve with well-known scoring rules (e.g., quadratic scoring rule)?
- We answer both at once, by considering approximate robustness.

Additional questions

- What if the cost function is not known to the experimenter?
- What can we achieve with well-known scoring rules (e.g., quadratic scoring rule)?
- We answer both at once, by considering approximate robustness.

Approximate robustness

- The experimenter cannot calibrate exactly the cost function.
- She estimates the cost function with a probability distribution.
- A scoring rule is (ε, δ)-robust if it elicits a belief within ε from the prior with probability at least $1 - \delta$.

Theorem

Assume that costs are posterior-separable almost surely, and consider arbitrary $\varepsilon > 0$ and $\delta > 0$. Then, for every proper scoring rule ϕ, there is some $\lambda \in (0, 1)$ such that $\lambda\phi$ is (ε, δ)-robust.

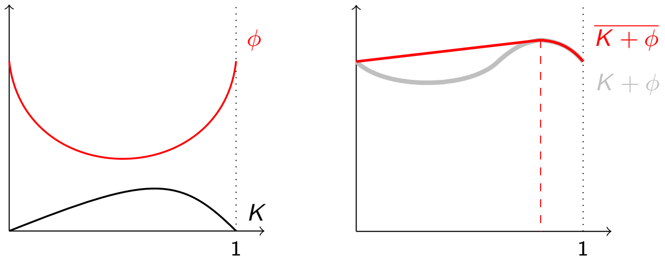
Approximate robustness

- The experimenter cannot calibrate exactly the cost function.
- She estimates the cost function with a probability distribution.
- A scoring rule is (ε, δ)-robust if it elicits a belief within ε from the prior with probability at least $1 - \delta$.

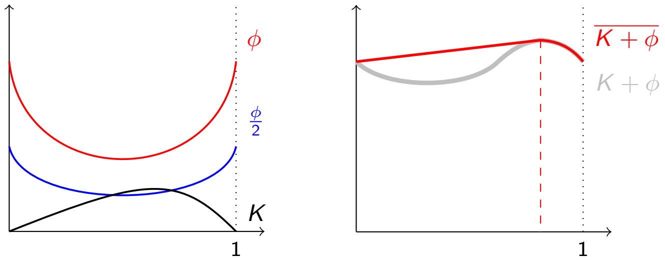
Theorem

Assume that costs are posterior-separable almost surely, and consider arbitrary $\varepsilon > 0$ and $\delta > 0$. Then, for every proper scoring rule ϕ, there is some $\lambda \in (0, 1)$ such that $\lambda\phi$ is (ε, δ)-robust.

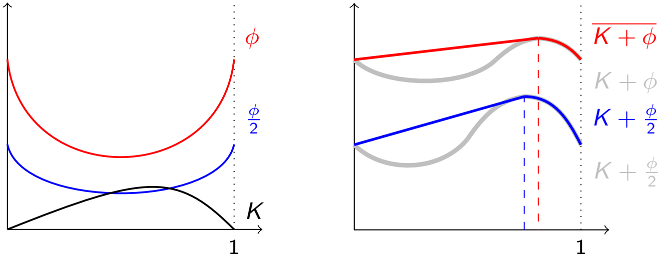
Approximate robustness intuitively



Approximate robustness intuitively

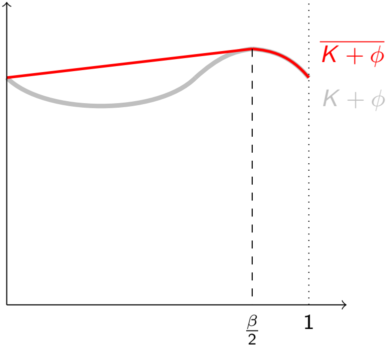


Approximate robustness intuitively



Approximately robust QSR

- **Quadratic scoring rule:** $\phi(\mu) = \alpha - \beta\mu(1-\mu)$
- **Costs:** $K(\mu) = \mu - \mu^3$ with probability 1.
- For arbitrary $\varepsilon > 0$, take $\beta = 2\varepsilon$ and the scoring rule will elicit a belief within ε from the prior (here we can even take $\delta = 0$).



Approximately robust QSR with uncertain entropic costs

- **Quadratic scoring rule:** $\phi(\mu) = \alpha - \beta\mu(1-\mu)$
- **Entropic costs:** $K(\mu) = -\kappa\big(\mu \log \mu + (1-\mu)\log(1-\mu)\big)$ with κ uniformly distributed in $[0,1]$
- If $\delta > 0$, then ϕ with $\beta \leq 2\delta$ will elicit the prior with probability at least $1-\delta$.

Roadmap

Cost of attention

Proposition

The cost function is **posterior-separable** *if and only if:*

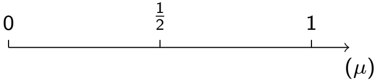
(C_1) NORMALIZATION: $C(\hat{\mu}) = 0$ *for all* $\mu \in [0, 1]$

(C_2) ATTENTION IS COSTLY: $C(\pi) > 0$ *for all* $\pi \in \hat{\Pi}(\mu) := \Pi(\mu) \setminus \{\hat{\mu}\}$ *and all* $\mu \in [0, 1]$

(C_3) DYNAMIC CONSISTENCY: *For all* $\pi \in \Pi(\mu)$ *and all* $\mu \in [0, 1]$,

$$C(\pi^*_\mu) = C(\pi) + \mathbb{E}_\pi(C \circ \pi^*) \quad (1)$$

- **Interpretation:** New information is always costly and the order of information does not matter (additive separability)



Cost of attention

Proposition

The cost function is **posterior-separable** *if and only if:*

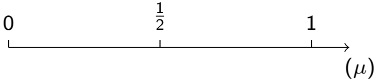
(C_1) NORMALIZATION: $C(\hat{\mu}) = 0$ *for all* $\mu \in [0, 1]$

(C_2) ATTENTION IS COSTLY: $C(\pi) > 0$ *for all* $\pi \in \hat{\Pi}(\mu) := \Pi(\mu) \setminus \{\hat{\mu}\}$ *and all* $\mu \in [0, 1]$

(C_3) DYNAMIC CONSISTENCY: *For all* $\pi \in \Pi(\mu)$ *and all* $\mu \in [0, 1]$,

$$C(\pi^*_\mu) = C(\pi) + \mathbb{E}_\pi(C \circ \pi^*) \qquad (1)$$

- **Interpretation:** New information is always costly and the order of information does not matter (additive separability)



Cost of attention

Proposition

The cost function is **posterior-separable** *if and only if:*

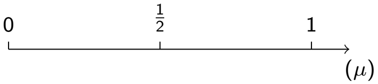
(C_1) NORMALIZATION: $C(\hat{\mu}) = 0$ *for all* $\mu \in [0,1]$

(C_2) ATTENTION IS COSTLY: $C(\pi) > 0$ *for all* $\pi \in \hat{\Pi}(\mu) := \Pi(\mu) \setminus \{\hat{\mu}\}$ *and all* $\mu \in [0,1]$

(C_3) DYNAMIC CONSISTENCY: *For all* $\pi \in \Pi(\mu)$ *and all* $\mu \in [0,1]$,

$$C(\pi^*_\mu) = C(\pi) + \mathbb{E}_\pi(C \circ \pi^*) \quad (1)$$

- **Interpretation:** New information is always costly and the order of information does not matter (additive separability)



Cost of attention

Proposition

The cost function is **posterior-separable** *if and only if:*

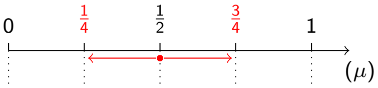
(C_1) NORMALIZATION: $C(\hat{\mu}) = 0$ *for all* $\mu \in [0,1]$

(C_2) ATTENTION IS COSTLY: $C(\pi) > 0$ *for all* $\pi \in \hat{\Pi}(\mu) := \Pi(\mu) \setminus \{\hat{\mu}\}$ *and all* $\mu \in [0,1]$

(C_3) DYNAMIC CONSISTENCY: *For all* $\pi \in \Pi(\mu)$ *and all* $\mu \in [0,1]$,

$$C(\pi^*_\mu) = C(\pi) + \mathbb{E}_\pi(C \circ \pi^*) \qquad (1)$$

- **Interpretation:** New information is always costly and the order of information does not matter (additive separability)



Cost of attention

Proposition

The cost function is **posterior-separable** *if and only if:*

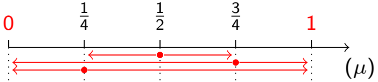
(C_1) NORMALIZATION: $C(\hat{\mu}) = 0$ *for all* $\mu \in [0, 1]$

(C_2) ATTENTION IS COSTLY: $C(\pi) > 0$ *for all* $\pi \in \hat{\Pi}(\mu) := \Pi(\mu) \setminus \{\hat{\mu}\}$ *and all* $\mu \in [0, 1]$

(C_3) DYNAMIC CONSISTENCY: *For all* $\pi \in \Pi(\mu)$ *and all* $\mu \in [0, 1]$,

$$C(\pi^*_\mu) = C(\pi) + \mathbb{E}_\pi(C \circ \pi^*) \tag{1}$$

- **Interpretation:** New information is always costly and the order of information does not matter (additive separability)



Cost of attention

Proposition

The cost function is **posterior-separable** *if and only if:*

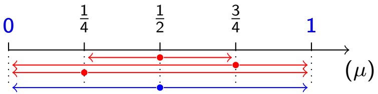
(C_1) NORMALIZATION: $C(\hat{\mu}) = 0$ *for all* $\mu \in [0,1]$

(C_2) ATTENTION IS COSTLY: $C(\pi) > 0$ *for all* $\pi \in \hat{\Pi}(\mu) := \Pi(\mu) \setminus \{\hat{\mu}\}$ *and all* $\mu \in [0,1]$

(C_3) DYNAMIC CONSISTENCY: *For all* $\pi \in \Pi(\mu)$ *and all* $\mu \in [0,1]$,

$$C(\pi^*_\mu) = C(\pi) + \mathbb{E}_\pi(C \circ \pi^*) \quad (1)$$

- **Interpretation:** New information is always costly and the order of information does not matter (additive separability)



Roadmap

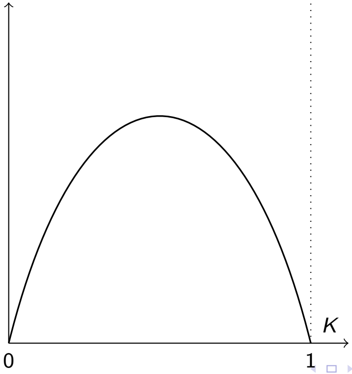
Exact robustness

Theorem

If the cost function is posterior-separable, there exists a robust scoring rule.

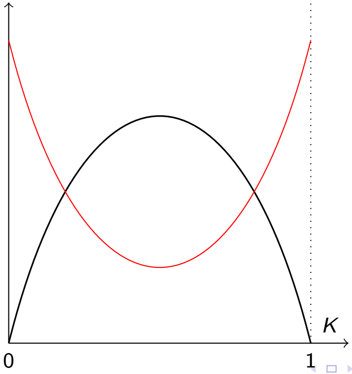
Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



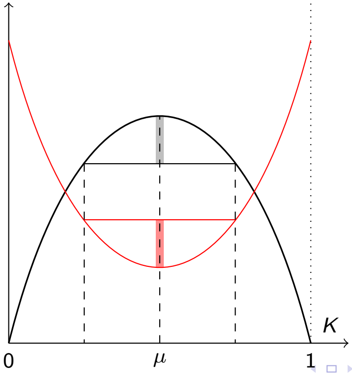
Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



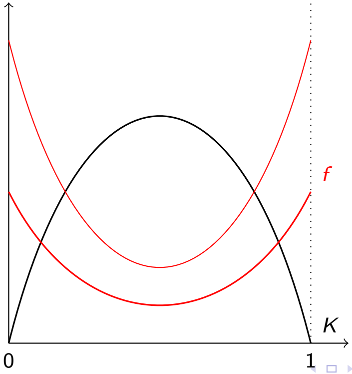
Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



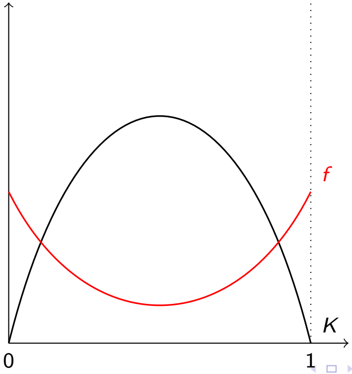
Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



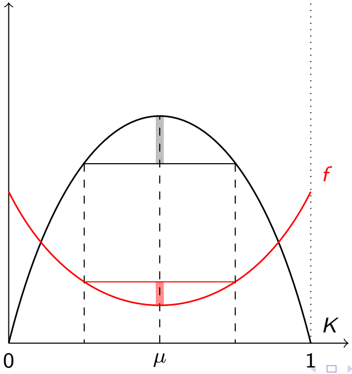
Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



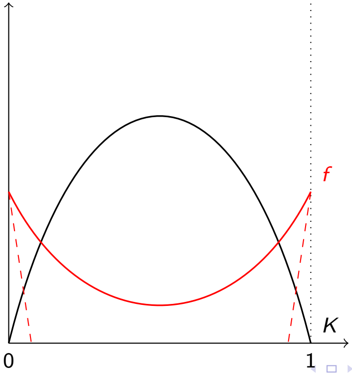
Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



Graphical sketch of the proof

1. Start with K and then take candidate scoring rule $a - K$.
2. Benefits equal costs.
3. Take $f := b(a - K)$ for some $b \in (0, 1)$.
4. Costs offset benefits ($K + f$ strictly concave).
5. Question remaining: is f subdifferentiable?



Sketch of the proof: boundary problem

Lemma

For every strictly convex function $f : [0,1] \to \mathbb{R}$, there exists some strictly convex $\phi : [0,1] \to \mathbb{R}$ such that

1. *$f - \phi$ is (weakly) convex ("ϕ is less convex than f")*
2. *$\partial\phi(\nu) \neq \emptyset$ for all $\nu \in \{0,1\}$ ("ϕ is subdifferentiable")*

- By (1), ϕ yields even weaker benefits than f.
- By (2), ϕ is a well-defined scoring rule.
- Hence, ϕ is a robust scoring rule (QED).

Sketch of the proof: boundary problem

Lemma

For every strictly convex function $f : [0,1] \to \mathbb{R}$, there exists some strictly convex $\phi : [0,1] \to \mathbb{R}$ such that

1. *$f - \phi$ is (weakly) convex ("ϕ is less convex than f")*
2. *$\partial\phi(\nu) \neq \emptyset$ for all $\nu \in \{0,1\}$ ("ϕ is subdifferentiable")*

- By (1), ϕ yields even weaker benefits than f.
- By (2), ϕ is a well-defined scoring rule.
- Hence, ϕ is a robust scoring rule (QED).

Sketch of the proof: boundary problem

Lemma

For every strictly convex function $f : [0,1] \to \mathbb{R}$, there exists some strictly convex $\phi : [0,1] \to \mathbb{R}$ such that

1. *$f - \phi$ is (weakly) convex ("ϕ is less convex than f")*
2. *$\partial\phi(\nu) \neq \emptyset$ for all $\nu \in \{0,1\}$ ("ϕ is subdifferentiable")*

- By (1), ϕ yields even weaker benefits than f.
- By (2), ϕ is a well-defined scoring rule.
- Hence, ϕ is a robust scoring rule (QED).

Sketch of the proof: boundary problem

Lemma

For every strictly convex function $f : [0,1] \to \mathbb{R}$, there exists some strictly convex $\phi : [0,1] \to \mathbb{R}$ such that

1. *$f - \phi$ is (weakly) convex ("ϕ is less convex than f")*
2. *$\partial\phi(\nu) \neq \emptyset$ for all $\nu \in \{0,1\}$ ("ϕ is subdifferentiable")*

- By (1), ϕ yields even weaker benefits than f.
- By (2), ϕ is a well-defined scoring rule.
- Hence, ϕ is a robust scoring rule (QED).

Thanks for listening!!!