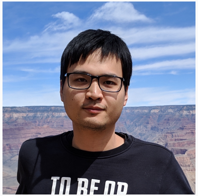
Mechanism Design for the Classroom (Optimization of Scoring Rules)

Jason Hartline

SLMath – Sept. 12, 2023

Northwestern University (visiting Stanford 2023–2024)
hartline@northwestern.edu



Yingkai Li

Liren Shan

Yifan Wu

Since Vickrey '61:

- . . . 62 years of research on mechanism design . . .

Applications of Mechanism Design

Since Vickrey '61:

- . . . 62 years of research on mechanism design . . .

Question: do novel mechanisms from the extensive literature work in practice?

Applications of Mechanism Design

Since Vickrey '61:

- . . . 62 years of research on mechanism design . . .

Question: do novel mechanisms from the extensive literature work in practice?

Some Areas of Successful Application:

- school choice [Abdulkadiroğlu, Sönmez AER'03] [. . .]
- matching doctors to hospitals [Roth, Peranson AER'99] [. . .]
- kidney exchange [Roth, Sönmez, Ünver QJE'04] [. . .]
- online advertising [Varian IJOR'07] [Edelman, Ostrovsky, Schwarz AER'07] [Edelman, Ostrovsky EC'11]
- spectrum auctions [Leyton-Brown, Milgrom, Segal PNAS'17] [. . .]

Since Vickrey '61:

- . . . 62 years of research on mechanism design . . .

Question: do novel mechanisms from the extensive literature work in practice?

Some Areas of Successful Application:

- school choice [Abdulkadiroğlu, Sönmez AER'03] [. . .]
- matching doctors to hospitals [Roth, Peranson AER'99] [. . .]
- kidney exchange [Roth, Sönmez, Ünver QJE'04] [. . .]
- online advertising [Varian IJOR'07] [Edelman, Ostrovsky, Schwarz AER'07] [Edelman, Ostrovsky EC'11]
- spectrum auctions [Leyton-Brown, Milgrom, Segal PNAS'17] [. . .]

Challenge: to test mechanism in practice, need strategic data for that mechanism!

Applications of Mechanism Design

Since Vickrey '61:

- ...62 years of research on mechanism design ...

Question: do novel mechanisms from the extensive literature work in practice?

Some Areas of Successful Application:

- school choice [Abdulkadiroğlu, Sönmez AER'03] [...]
- matching doctors to hospitals [Roth, Peranson AER'99] [...]
- kidney exchange [Roth, Sönmez, Ünver QJE'04] [...]
- online advertising [Varian IJOR'07] [Edelman, Ostrovsky, Schwarz AER'07] [Edelman, Ostrovsky EC'11]
- spectrum auctions [Leyton-Brown, Milgrom, Segal PNAS'17] [...]

Challenge: to test mechanism in practice, need strategic data for that mechanism! Very difficult!

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

- grading randomized exams: ex post fairness?

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

- grading randomized exams: ex post fairness?
- grading with partial credit: incentivizing precise answers?

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

- grading randomized exams: ex post fairness?
- grading with partial credit: incentivizing precise answers?
- group projects: incentivizing teamwork?

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

- grading randomized exams: ex post fairness?
- grading with partial credit: incentivizing precise answers?
- group projects: incentivizing teamwork?
- peer grading: incentives for accurate peer reviews?

Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

- grading randomized exams: ex post fairness? [Chen, Hartline, Zoeter FORC'23]
- grading with partial credit: incentivizing precise answers? [Chen, Hartline, Zoeter]
- group projects: incentivizing teamwork?
- peer grading: incentives for accurate peer reviews? [Li, Hartline, Shan, Wu EC'22]

A Peer Grading Story

1. A peer grading platform (PeerPal).
2. Grading peer reviews with proper scoring rules is horrible!
3. (Quick fix: Manually grade the peer reviews.)
4. Optimization of scoring rules.
5. Fundamental Role of Scoring Rules

PeerPal

Peer grading system:

- Canvas plugin (`https://www.peerpal.io/`)
- used in "Intro to CS", "Intro to Algorithms", "Intro to Online Markets", "Mechanism Design", etc.

Peer grading system:

- Canvas plugin (`https://www.peerpal.io/`)
- used in "Intro to CS", "Intro to Algorithms", "Intro to Online Markets", "Mechanism Design", etc.

Main Algorithms:

- matching peers and TAs to submissions
- grading submissions from peer reviews
- grading peer reviews from TA reviews

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.

- reduces teacher grading.

- promptness of feedback.

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.
  (learn material: 60% agree; learn to write better: 55% agree
  (worse students agree more: A: 52%; B: 54%; C: 75%; D: 80%)
- reduces teacher grading.
- promptness of feedback.

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.
  (learn material: 60% agree; learn to write better: 55% agree
  (worse students agree more: A: 52%; B: 54%; C: 75%; D: 80%)
- reduces teacher grading.
  (TAs graded 1/5 of student work.)
- promptness of feedback.

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.
  (learn material: 60% agree; learn to write better: 55% agree
  (worse students agree more: A: 52%; B: 54%; C: 75%; D: 80%)
- reduces teacher grading.
  (TAs graded 1/5 of student work.)
- promptness of feedback.
  (peer review feedback in 3 days, grades in 5 days; versus 2 weeks)

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.
  (learn material: 60% agree; learn to write better: 55% agree
  (worse students agree more: A: 52%; B: 54%; C: 75%; D: 80%)
- reduces teacher grading.
  (TAs graded 1/5 of student work.)
- promptness of feedback.
  (peer review feedback in 3 days, grades in 5 days; versus 2 weeks)

Potential Disadvantages: Inaccurate grades, student unrest, . . .

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.
  (learn material: 60% agree; learn to write better: 55% agree
  (worse students agree more: A: 52%; B: 54%; C: 75%; D: 80%)
- reduces teacher grading.
  (TAs graded 1/5 of student work.)
- promptness of feedback.
  (peer review feedback in 3 days, grades in 5 days; versus 2 weeks)

Potential Disadvantages: Inaccurate grades, student unrest, . . .

(3.7% appeal rate; 1-6% strongly disagree with survey questions)

Advantages of Peer Grading

Advantages of Peer Grading: (observations from Intro to Algs)

- learning by reviewing.
  (learn material: 60% agree; learn to write better: 55% agree
  (worse students agree more: A: 52%; B: 54%; C: 75%; D: 80%)
- reduces teacher grading.
  (TAs graded 1/5 of student work.)
- promptness of feedback.
  (peer review feedback in 3 days, grades in 5 days; versus 2 weeks)

Potential Disadvantages: Inaccurate grades, student unrest, . . .

(3.7% appeal rate; 1-6% strongly disagree with survey questions)

Main Challenge: incentivizing accurate peer reviews.

(i.e., "grading the grading")

Grading Peer Reviews

Example Scenario:

- 100 students
- submit homeworks in pairs $\Rightarrow$ 50 submissions.
- each review three submissions $\Rightarrow$ 300 peer reviews.
- need to grade: 50 submissions, 300 peer reviews.

Grading Peer Reviews

Example Scenario:

- 100 students
- submit homeworks in pairs $\Rightarrow$ 50 submissions.
- each review three submissions $\Rightarrow$ 300 peer reviews.
- need to grade: 50 submissions, 300 peer reviews.

Approach:

1. pick 10 submissions for TA to review.
2. assign each peer 1 of these 10 submissions at random to review.
3. assign each peer 2 of remaining 40 submissions at random to review.
4. grade Step 2 peer reviews against TA review. (don't grade other reviews)

Grading Peer Reviews

Example Scenario:

- 100 students
- submit homeworks in pairs $\Rightarrow$ 50 submissions.
- each review three submissions $\Rightarrow$ 300 peer reviews.
- need to grade: 50 submissions, 300 peer reviews.

Approach:

1. pick 10 submissions for TA to review.
2. assign each peer 1 of these 10 submissions at random to review.
3. assign each peer 2 of remaining 40 submissions at random to review.
4. grade Step 2 peer reviews against TA review. (don't grade other reviews)

Remaining challenge: grading peer reviews from TA review.

Grading Peer Reviews

Example Scenario:

- 100 students
- submit homeworks in pairs $\Rightarrow$ 50 submissions.
- each review three submissions $\Rightarrow$ 300 peer reviews.
- need to grade: 50 submissions, 300 peer reviews.

Approach:

1. pick 10 submissions for TA to review.
2. assign each peer 1 of these 10 submissions at random to review.
3. assign each peer 2 of remaining 40 submissions at random to review.
4. grade Step 2 peer reviews against TA review. (don't grade other reviews)

Remaining challenge: grading peer reviews from TA review.

Idea: use proper scoring rule! [McCarthy PNAS'56] [Savage JASA'71] [Gneiting, Raftery JASA'07] [. . .].

A Peer Grading Story

1. A peer grading platform (PeerPal).
2. Grading peer reviews with proper scoring rules is horrible!
3. (Quick fix: Manually grade the peer reviews.)
4. Optimization of scoring rules.
5. Fundamental Role of Scoring Rules

Grading Review with Proper Scoring Rule

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$

Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Example

Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Example

- peer report $r = 0.8$

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0,1]$ (truth)
- Peer belief $\mu \in [0,1]$; peer report $r \in [0,1]$
- quadratic scoring rule: $S(r,\theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r,\theta) = 1 - 0.5^2 = 0.75$

Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Theorem
Reporting $r = \mu$ is optimal for peer.

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r, \theta) = 1 - 0.5^2 = 0.75$

Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Theorem
Reporting $r = \mu$ is optimal for peer.

Proof.

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r, \theta) = 1 - 0.5^2 = 0.75$

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

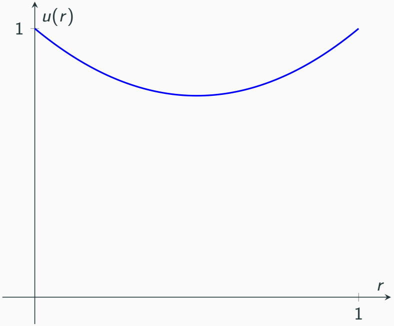
Theorem
Reporting $r = \mu$ is optimal for peer.

Proof.

- let $u(r) = 1 - r + r^2$

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r, \theta) = 1 - 0.5^2 = 0.75$



Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Theorem
Reporting $r = \mu$ is optimal for peer.

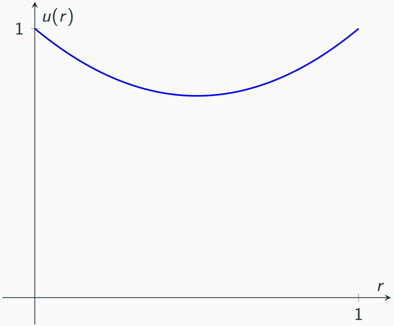
Proof.

- let $u(r) = 1 - r + r^2$
- algebra $\Rightarrow$ can rewrite as:

$$S(r, \theta) = u(r) + u'(r)\,(\theta - r) + \kappa(\theta).$$

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r, \theta) = 1 - 0.5^2 = 0.75$



Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0,1]$ (truth)
- Peer belief $\mu \in [0,1]$; peer report $r \in [0,1]$
- quadratic scoring rule: $S(r,\theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

Theorem
Reporting $r = \mu$ *is optimal for peer.*

Proof.

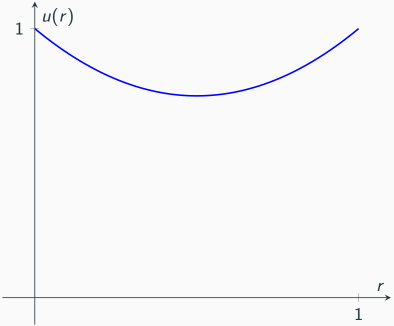
- let $u(r) = 1 - r + r^2$
- algebra $\Rightarrow$ can rewrite as:

$$S(r,\theta) = u(r) + u'(r)\,(\theta - r) + \kappa(\theta).$$

- report cannot affect κ (so ignore it)

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r,\theta) = 1 - 0.5^2 = 0.75$



Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0, 1]$ (truth)
- Peer belief $\mu \in [0, 1]$; peer report $r \in [0, 1]$
- quadratic scoring rule: $S(r, \theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

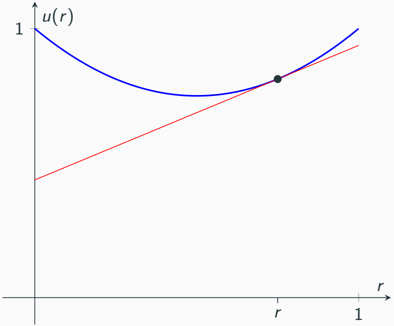
Theorem
Reporting $r = \mu$ is optimal for peer.

Proof.

- let $u(r) = 1 - r + r^2$
- algebra $\Rightarrow$ can rewrite as:
$$S(r, \theta) = u(r) + u'(r)\,(\theta - r) + \kappa(\theta).$$
- report cannot affect κ (so ignore it)
- let supporting tangent at r be:
$$h_r(\theta) = u(r) + u'(r)(\theta - r)$$

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r, \theta) = 1 - 0.5^2 = 0.75$



Proper Scoring Rules

Grading Review with Proper Scoring Rule

- TA score $\theta \in [0,1]$ (truth)
- Peer belief $\mu \in [0,1]$; peer report $r \in [0,1]$
- quadratic scoring rule: $S(r,\theta) = 1 - (\theta - r)^2$
- (if multiple rubric elements, average across rubric)

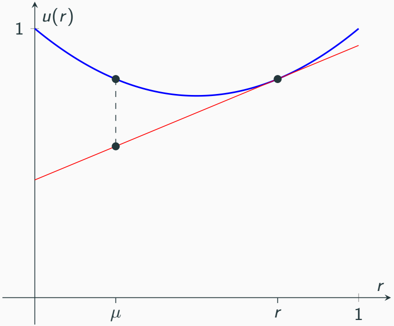
Theorem
Reporting $r = \mu$ is optimal for peer.

Proof.

- let $u(r) = 1 - r + r^2$
- algebra $\Rightarrow$ can rewrite as:
$$S(r,\theta) = u(r) + u'(r)\,(\theta - r) + \kappa(\theta).$$
- report cannot affect κ (so ignore it)
- let supporting tangent at r be:
$$h_r(\theta) = u(r) + u'(r)(\theta - r)$$
- loss from report r at belief μ: $u(\mu) - h_r(\mu)$. $\square$

Example

- peer report $r = 0.8$
- TA report $\theta = 0.3$
- score $S(r,\theta) = 1 - 0.5^2 = 0.75$



The Lazy Peer Strategy

Standard Scoring Rules are Horrible

The Lazy Peer Strategy

- assume: TA grade $\theta \in [0.6, 1]$

The Lazy Peer Strategy

- assume: TA grade $\theta \in [0.6, 1]$
- strategy: always report $r = 0.8$

Standard Scoring Rules are Horrible

The Lazy Peer Strategy

- assume: TA grade $\theta \in [0.6, 1]$
- strategy: always report $r = 0.8$
- $S(r, \theta) \geq 1 - (0.2)^2 = 0.96$

Standard Scoring Rules are Horrible

The Lazy Peer Strategy

- assume: TA grade $\theta \in [0.6, 1]$
- strategy: always report $r = 0.8$
- $S(r, \theta) \geq 1 - (0.2)^2 = 0.96$

Result
Very little incentive for effort!

A Peer Grading Story

1. A peer grading platform (PeerPal).
2. Grading peer reviews with proper scoring rules is horrible!
3. (Quick fix: Manually grade the peer reviews.)
4. Optimization of scoring rules.
5. Fundamental Role of Scoring Rules

Review Grading By Hand

Submission 42

⋮

contents of submission

⋮

	Peer 1	Peer 2	Peer 3	TA Score	TA Comment
Algorithm	8*	9*	10	9	good solution ...
Correctness	5*	7*	10	6	missing base case ...
Clarity	8*	8*	10	8	easy to follow ...
Quantitative	9	10	5		
Qualitative	8	8	0		
Feedback	see TA review	see TA review	must provide detailed review		

1. A peer grading platform (PeerPal).
2. Grading peer reviews with proper scoring rules is horrible!
3. (Quick fix: Manually grade the peer reviews.)
4. Optimization of scoring rules.
5. Fundamental Role of Scoring Rules

Summary: Optimization of Scoring Rules

Optimal Scoring Rule for Incentivizing Binary Effort

- peers choose effort or no effort
- maximize: difference in score for effort vs no effort
- subject to: proper and bounded scoring rule.

Summary: Optimization of Scoring Rules

Optimal Scoring Rule for Incentivizing Binary Effort

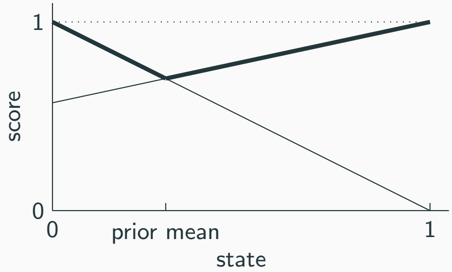
$$\begin{aligned} \max_{\text{scoring rule}} \ & \mathbf{E}_{\text{state, belief with effort}}[\text{score with effort} - \text{score without effort}] \\ \text{s.t.} \ & \text{scoring rule is proper (optimal to truthfully report belief)} \\ & \text{scoring rule is bounded} \end{aligned}$$

Optimal Scoring Rule for Incentivizing Binary Effort

$$\max_{\text{scoring rule}} \mathbf{E}_{\text{state, belief with effort}}[\text{score with effort} - \text{score without effort}]$$
$$\text{s.t. scoring rule is proper (optimal to truthfully report belief)}$$
$$\text{scoring rule is bounded}$$

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state



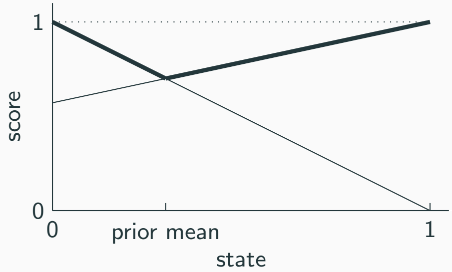
Optimal Scoring Rule for Incentivizing Binary Effort

$$\max_{\text{scoring rule}} \mathbf{E}_{\text{state, belief with effort}}[\text{score with effort} - \text{score without effort}]$$
$$\text{s.t. scoring rule is proper (optimal to truthfully report belief)}$$
$$\text{scoring rule is bounded}$$

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

(standard scoring rules like quadratic not approx optimal)



Summary: Optimization of Scoring Rules

Optimal Scoring Rule for Incentivizing Binary Effort

$$\max_{\text{scoring rule}} \ \mathbf{E}_{\text{state, belief with effort}}[\text{score with effort} - \text{score without effort}]$$
$$\text{s.t. scoring rule is proper (optimal to truthfully report belief)}$$
$$\text{scoring rule is bounded}$$

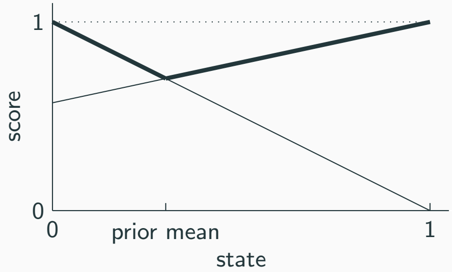
Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

(standard scoring rules like quadratic not approx optimal)

Theorem
approximately optimal multi-dimensional scoring rule:

maximum over optimal separate scoring rules



Optimal Scoring Rule for Incentivizing Binary Effort

$$\max_{\text{scoring rule}} \mathbf{E}_{\text{state, belief with effort}}[\text{score with effort} - \text{score without effort}]$$
$$\text{s.t. scoring rule is proper (optimal to truthfully report belief)}$$
$$\text{scoring rule is bounded}$$

Theorem
optimal single-dimensional scoring rule:

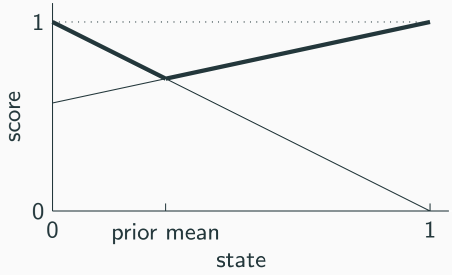
choose side of prior mean, score linear in state

(standard scoring rules like quadratic not approx optimal)

Theorem
approximately optimal multi-dimensional scoring rule:

maximum over optimal separate scoring rules

(average of separate scoring rules not approx optimal)



Single-dimensional Optimal Scoring Rules

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

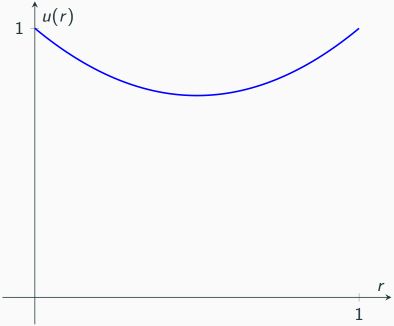
Proof.

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u



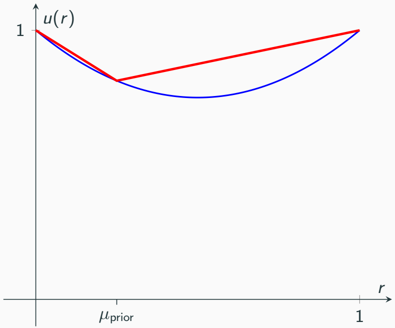
Single-dimensional Optimal Scoring Rules

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}



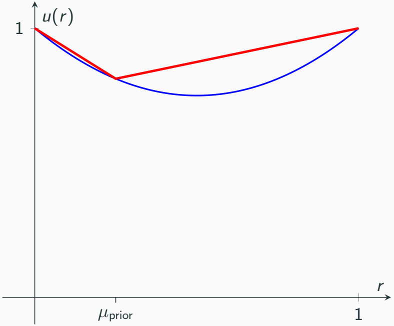
Single-dimensional Optimal Scoring Rules

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}
- objective $\mathbf{E}[u(\mu_{\text{posterior}}) - u(\mu_{\text{prior}})]$ weakly increased:



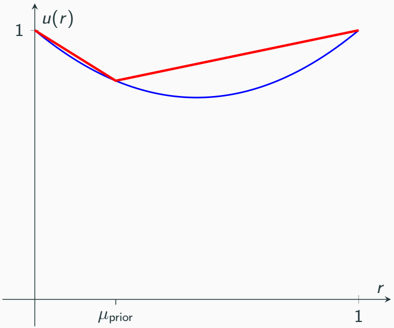
Single-dimensional Optimal Scoring Rules

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}
- objective $\mathbf{E}[u(\mu_{\text{posterior}}) - u(\mu_{\text{prior}})]$ weakly increased:
  - $u(\mu_{\text{posterior}})$ weakly increased.



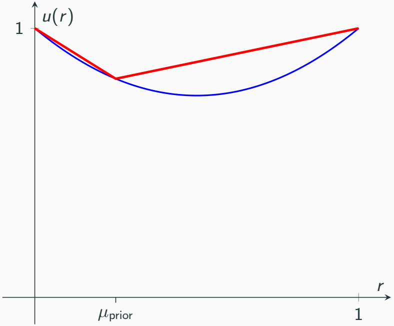
Single-dimensional Optimal Scoring Rules

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}
- objective $\mathbf{E}[u(\mu_{\text{posterior}}) - u(\mu_{\text{prior}})]$ weakly increased:
    - $u(\mu_{\text{posterior}})$ weakly increased.
    - $u(\mu_{\text{prior}})$ is unchanged.



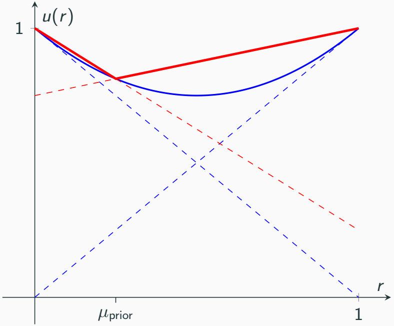
Single-dimensional Optimal Scoring Rules

Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}
- objective $\mathbf{E}[u(\mu_{\text{posterior}}) - u(\mu_{\text{prior}})]$ weakly increased:
  - $u(\mu_{\text{posterior}})$ weakly increased.
  - $u(\mu_{\text{prior}})$ is unchanged.
- score for extremal reports weakly less extreme

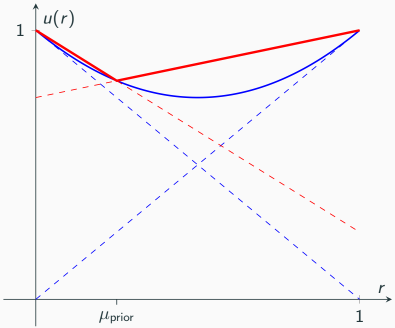


Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}
- objective $\mathbf{E}[u(\mu_{\text{posterior}}) - u(\mu_{\text{prior}})]$ weakly increased:
  - $u(\mu_{\text{posterior}})$ weakly increased.
  - $u(\mu_{\text{prior}})$ is unchanged.
- score for extremal reports weakly less extreme
  - still ex post bounded.

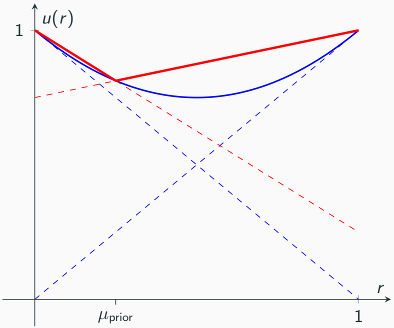


Theorem
optimal single-dimensional scoring rule:

choose side of prior mean, score linear in state

Proof.

- consider ex post bounded scoring rule defined by convex u
- replace $u(r)$ with V-shape at μ_{prior}
- objective $\mathbf{E}[u(\mu_{\text{posterior}}) - u(\mu_{\text{prior}})]$ weakly increased:
  - $u(\mu_{\text{posterior}})$ weakly increased.
  - $u(\mu_{\text{prior}})$ is unchanged.
- score for extremal reports weakly less extreme
  - still ex post bounded. □



Related Work:

- characterizing scoring rules:
  - eliciting full distribution [McCarthy '56; Gneiting, Raftery '07]
  - eliciting the mean [Abernethy, Frongillo '12]
  - set of elicitable properties (e.g., variance is not directly elicitable) [Lambert '11]

Related Work:

- characterizing scoring rules:
  - eliciting full distribution [McCarthy '56; Gneiting, Raftery '07]
  - eliciting the mean [Abernethy, Frongillo '12]
  - set of elicitable properties (e.g., variance is not directly elicitable) [Lambert '11]
- maximize effort with quadratic scoring rules [Osband '89]

Related Work

Related Work:

- characterizing scoring rules:
  - eliciting full distribution [McCarthy '56; Gneiting, Raftery '07]
  - eliciting the mean [Abernethy, Frongillo '12]
  - set of elicitable properties (e.g., variance is not directly elicitable) [Lambert '11]
- maximize effort with quadratic scoring rules [Osband '89]
- maximize effort in a binary state model with costly samples [Neyman, Noarov, Weinberg '21]

Related Work:

- characterizing scoring rules:
  - eliciting full distribution [McCarthy '56; Gneiting, Raftery '07]
  - eliciting the mean [Abernethy, Frongillo '12]
  - set of elicitable properties (e.g., variance is not directly elicitable) [Lambert '11]
- maximize effort with quadratic scoring rules [Osband '89]
- maximize effort in a binary state model with costly samples [Neyman, Noarov, Weinberg '21]
- framework adopted by follow-up works:
  - optimizing max-min objective without knowledge about prior and signal [Chen and Yu '21]
  - optimization of peer prediction mechanisms [Kong '21]
  - bounded expected score [Papireddygari, Waggoner '22]
  - maximizing effort under multi-dimensional effort model [Hartline, Li, Shan, Wu '23]
  - benchmark for visualization experiments [Wu, Guo, Mamakos, Hartline, Hullman '23]

Related Work:

- characterizing scoring rules:
  - eliciting full distribution [McCarthy '56; Gneiting, Raftery '07]
  - eliciting the mean [Abernethy, Frongillo '12]
  - set of elicitable properties (e.g., variance is not directly elicitable) [Lambert '11]
- maximize effort with quadratic scoring rules [Osband '89]
- maximize effort in a binary state model with costly samples [Neyman, Noarov, Weinberg '21]
- framework adopted by follow-up works:
  - optimizing max-min objective without knowledge about prior and signal [Chen and Yu '21]
  - optimization of peer prediction mechanisms [Kong '21]
  - bounded expected score [Papireddygari, Waggoner '22]
  - maximizing effort under multi-dimensional effort model [Hartline, Li, Shan, Wu '23]
  - benchmark for visualization experiments [Wu, Guo, Mamakos, Hartline, Hullman '23]
- ex post value of information [Frankel, Kamenica '19]

A Peer Grading Story

1. A peer grading platform (PeerPal).

2. Grading peer reviews with proper scoring rules is horrible!

3. (Quick fix: Manually grade the peer reviews.)

4. Optimization of scoring rules.

5. Fundamental Role of Scoring Rules

A Value of Data (via "Revelation Principle")

analysis of dataset $\longrightarrow$ decision optimization $\longrightarrow$ payoff from decision

A Value of Data (via "Revelation Principle")

analysis of dataset $\longrightarrow$ $\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}$

proper scoring rule

A Value of Data (via "Revelation Principle")

analysis of dataset $\longrightarrow$ $\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}$

proper scoring rule

Interpretations

A Value of Data (via "Revelation Principle")

$$\text{analysis of dataset} \longrightarrow \underbrace{\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}}_{\text{proper scoring rule}}$$

Interpretations

- scoring rules are fundamental for understanding good data analyses

A Value of Data (via "Revelation Principle")

$$\text{analysis of dataset} \longrightarrow \underbrace{\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}}_{\text{proper scoring rule}}$$

Interpretations

- scoring rules are fundamental for understanding good data analyses
- optimal scoring rules for binary effort $\Rightarrow$ setting-independent value of dataset

A Value of Data (via "Revelation Principle")

$$\text{analysis of dataset} \longrightarrow \underbrace{\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}}_{\text{proper scoring rule}}$$

Interpretations

- scoring rules are fundamental for understanding good data analyses
- optimal scoring rules for binary effort $\Rightarrow$ setting-independent value of dataset

Example (Rational Agent Framework for Data Visualization)

[Wu, Guo, Mamakos, Hartline, Hullman VIS'23]

A Value of Data (via "Revelation Principle")

$$\text{analysis of dataset} \longrightarrow \underbrace{\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}}_{\text{proper scoring rule}}$$

Interpretations

- scoring rules are fundamental for understanding good data analyses
- optimal scoring rules for binary effort $\Rightarrow$ setting-independent value of dataset

Example (Rational Agent Framework for Data Visualization)

- researcher shows behavioral subjects different visual stimuli.

[Wu, Guo, Mamakos, Hartline, Hullman VIS'23]

A Value of Data (via "Revelation Principle")

$$\text{analysis of dataset} \longrightarrow \underbrace{\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}}_{\text{proper scoring rule}}$$

Interpretations

- scoring rules are fundamental for understanding good data analyses
- optimal scoring rules for binary effort $\Rightarrow$ setting-independent value of dataset

Example (Rational Agent Framework for Data Visualization)

- researcher shows behavioral subjects different visual stimuli.
- measure performance in decision problem (a.k.a., scoring rule).

[Wu, Guo, Mamakos, Hartline, Hullman VIS'23]

A Value of Data (via "Revelation Principle")

$$\text{analysis of dataset} \longrightarrow \underbrace{\boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}}_{\text{proper scoring rule}}$$

Interpretations

- scoring rules are fundamental for understanding good data analyses
- optimal scoring rules for binary effort $\Rightarrow$ setting-independent value of dataset

Example (Rational Agent Framework for Data Visualization)

- researcher shows behavioral subjects different visual stimuli.
- measure performance in decision problem (a.k.a., scoring rule).
- benchmark against rational agent with and without stimuli.

[Wu, Guo, Mamakos, Hartline, Hullman VIS'23]

A Value of Data (via “Revelation Principle”)

$$\text{analysis of dataset} \longrightarrow \boxed{\text{decision optimization} \longrightarrow \text{payoff from decision}}$$
$$\text{proper scoring rule}$$

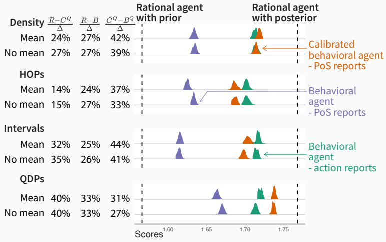
Interpretations

- scoring rules are fundamental for understanding good data analyses
- optimal scoring rules for binary effort $\Rightarrow$ setting-independent value of dataset

Example (Rational Agent Framework for Data Visualization)

- researcher shows behavioral subjects different visual stimuli.
- measure performance in decision problem (a.k.a., scoring rule).
- benchmark against rational agent with and without stimuli.

[Wu, Guo, Mamakos, Hartline, Hullman VIS’23]



Mechanism Design for the Classroom

The classroom as a market:

- students: agents
- instructor: principal
- syllabus: rules that map actions to grades
- student incentives: minimize work, maximize grade
- goal: minimize work, maximize learning, fairly assess

Basic Questions: What is best syllabus?

Examples:

- grading randomized exams: ex post fairness? [Chen, Hartline, Zoeter FORC'23]
- grading with partial credit: incentivizing precise answers? [Chen, Hartline, Zoeter]
- group projects: incentivizing teamwork?
- peer grading: incentives for accurate peer reviews? [Li, Hartline, Shan, Wu EC'22]