ExpEcon Methods: Incentive Compatible Belief Elicitation

ECON 8877

P.J. Healy

Updated 2023-11-01 at 17:41:12

Overview

- We often want to elicit the subject's belief about an event
    - Opponent's action in a game
    - Own absolute performance on a quiz/task
    - Performance in the top half
    - Guess the performance of someone else
    - Bayesian updating tests
- But there are many ways proposed to do this!
    - Quadratic scoring rule
    - Logarithmic scoring rule
    - Spherical scoring rule
    - Binarized scoring rule
    - BDM for probabilities
        - Auction framing
        - Two random variables framing
        - MPL framing

Our Framework

- Always specify your framework! Savage? Segal? AA?
- Savage: need entire $\succeq$ to learn beliefs
  - That's too many questions!
  - ...and requires probabilistic sophistication
- vNM/Segal: no subjective beliefs!
- AA: can compare against objective lotteries
  - Having "belief" p means I'm indifferent between:
    1. Getting \$$x$ if E occurs
    2. Getting \$$x$ with probability p
  - Call the indifference point $p(E, x)$
  - Stakes independence (analogue of P4):
    - $p(E, x) = p(E, y) = p(E) \ \forall x, y > 0$
    - Question: Which AA/Seo axioms give this?
    - Do we really even need this??

Setup

- Random variable $X : \Omega \to \mathbb{R}$
- Subject has belief $p(X = x)$ for each realization x
- Example: probability of event E
  - Let $X_E = 1$ if $\omega \in E$, $X_E = 0$ otherwise (indicator)
  - $p(E) := p(X_E = 1)$

Proportions vs. Probabilities

Application: What fraction of opponents chose Cooperate?
Two options:

1. What fraction of people chose C?
   - Call the true fraction $\rho \in [0, 1]$
   - Subject has a belief over *all* of $[0, 1]$
   - Their belief is an entire PDF/CDF!!
   - Later: we can elicit mean, median, mode, etc.
2. What's the probability a random opponent chose C?
   - Now the truth is either 0 or 1
   - Subject has a belief $p \in [0, 1]$
   - Here we just elicit a single probability

Scoring Rules

- Used to elicit $p(E)$
- Subject announces q
- State-contingent payment:
  1. $\$S(q, 1)$ if $X_E = 1$
  2. $\$S(q, 0)$ if $X_E = 0$
- True belief: p
- Expected payoff: $G(q|p) = pS(q, 1) + (1 - p)S(q, 0)$
- Scoring rule S is *proper* if

$$p \in \arg\max_q G(q|p)$$

  and *strictly proper* if

$$p = \arg\max_q G(q|p)$$

- Under risk-neutral EU, proper $\Rightarrow$ IC
- Let $G(p) = G(p|p)$ (used later)

Example: Quadratic Scoring Rule

The original scoring rule: Brier (1950)

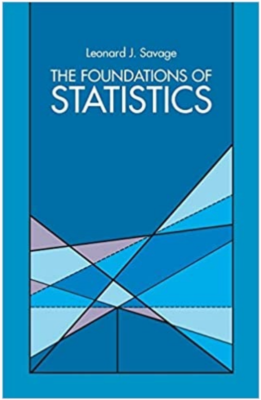
- $S(q, 1) = \$1 - \$(1 - q)^2$
- $S(q, 0) = \$1 - \$(0 - q)^2$
- General: $S(q, X_E) = 1 - (X_E - q)^2$

$$
\begin{aligned}
G(q|p) &= p[1 - (1 - q)^2] + (1 - p)[1 - (0 - q)^2] \\
&= -p(1 - q)^2 - (1 - p)q^2 \\
\frac{\partial G(q|p)}{\partial q} &= 2p(1 - q) - 2(1 - p)q = 0 \\
& p(1 - q) = (1 - p)q \\
& q^* = p
\end{aligned}
$$

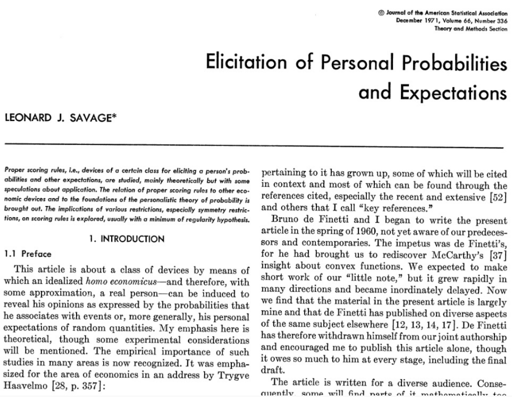
Can rescale it and it's still strictly proper:

- $S(q, 1) = \beta_1 - \alpha(1 - q)^2$
- $S(q, 0) = \beta_0 - \alpha(0 - q)^2$

Theory: Savage (1971)



(1954)



© Journal of the American Statistical Association
December 1971, Volume 66, Number 336
Theory and Methods Section

Elicitation of Personal Probabilities and Expectations

LEONARD J. SAVAGE*

Proper scoring rules, i.e., devices of a certain class for eliciting a person's probabilities and other expectations, are studied, mainly theoretically but with some speculations about application. The relation of proper scoring rules to other economic devices and to the foundations of the personalistic theory of probability is brought out. The implications of various restrictions, especially symmetry restrictions, on scoring rules is explored, usually with a minimum of regularity hypothesis.

1. INTRODUCTION

1.1 Preface

This article is about a class of devices by means of which an idealized *homo economicus*—and therefore, with some approximation, a real person—can be induced to reveal his opinions as expressed by the probabilities that he associates with events or, more generally, his personal expectations of random quantities. My emphasis here is theoretical, though some experimental considerations will be mentioned. The empirical importance of such studies in many areas is now recognized. It was emphasized for the area of economics in an address by Trygve Haavelmo [28, p. 357]:

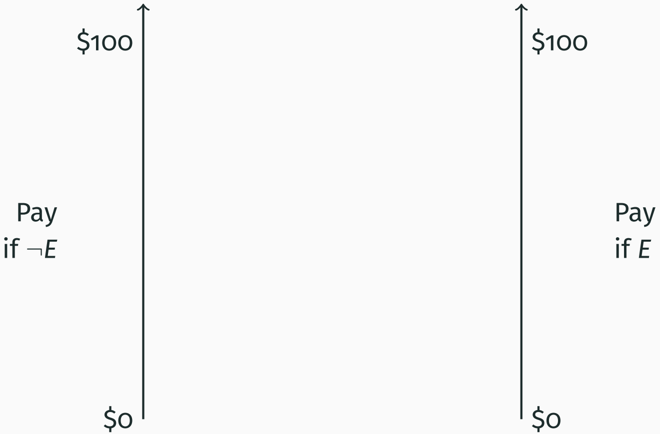
pertaining to it has grown up, some of which will be cited in context and most of which can be found through the references cited, especially the recent and extensive [52] and others that I call "key references."

Bruno de Finetti and I began to write the present article in the spring of 1960, not yet aware of our predecessors and contemporaries. The impetus was de Finetti's, for he had brought us to rediscover McCarthy's [37] insight about convex functions. We expected to make short work of our "little note," but it grew rapidly in many directions and became inordinately delayed. Now we find that the material in the present article is largely mine and that de Finetti has published on diverse aspects of the same subject elsewhere [12, 13, 14, 17]. De Finetti has therefore withdrawn himself from our joint authorship and encouraged me to publish this article alone, though it owes so much to him at every stage, including the final draft.

The article is written for a diverse audience. Conse-

(1971)

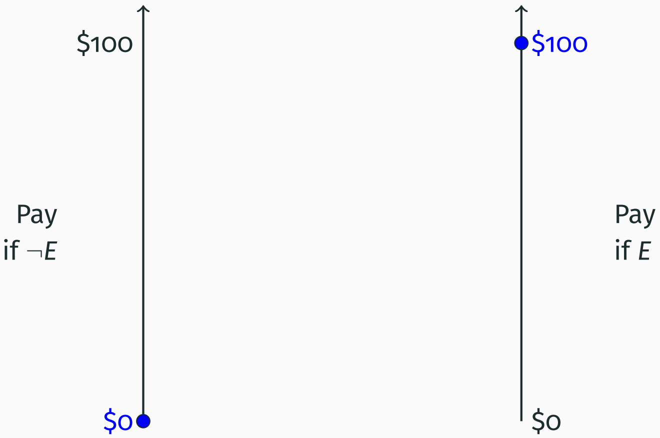
Theory: Savage (1971)



Want to know subject's $Pr(E)$ for some event E

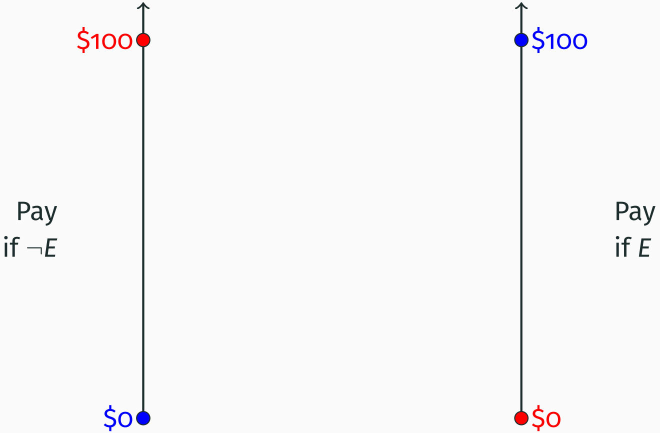
Pay using state-contingent payments ('bets')

Theory: Savage (1971)



Example: A \$100 bet on *E*

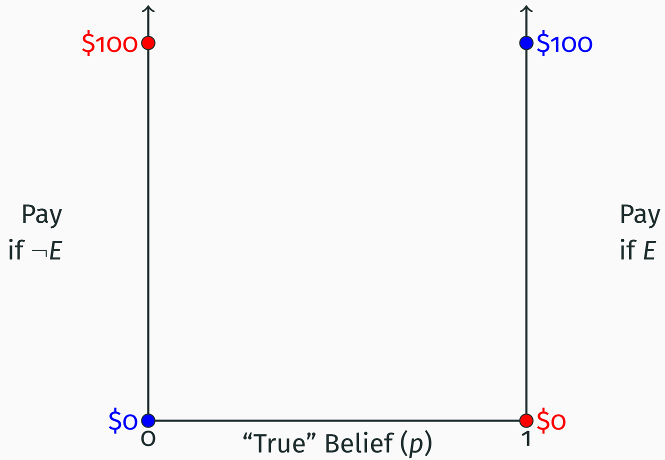
Theory: Savage (1971)



Example: A $100 bet on *E*

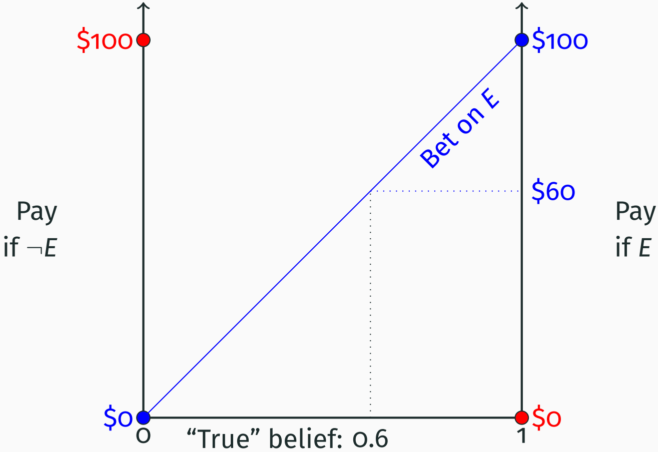
A $100 bet on ¬*E*

Theory: Savage (1971)



How you evaluate these depends on your "true" belief

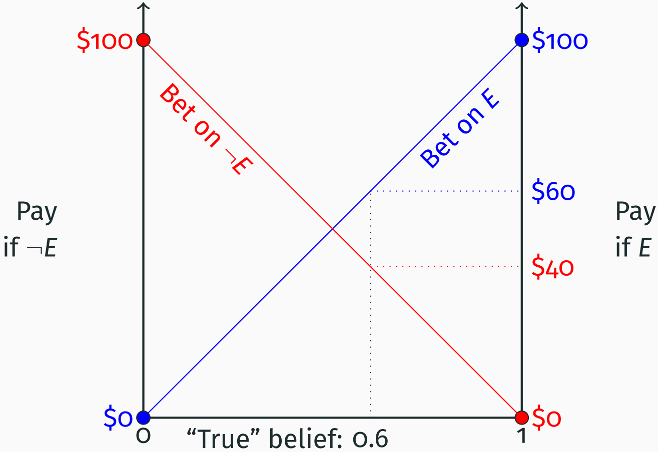
Theory: Savage (1971)



How you evaluate these depends on your "true" belief

Assume (for now) risk-neutral EU

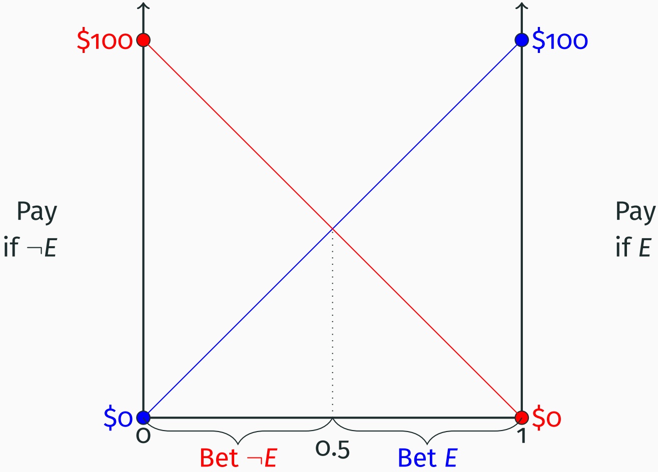
Theory: Savage (1971)



How you evaluate these depends on your "true" belief

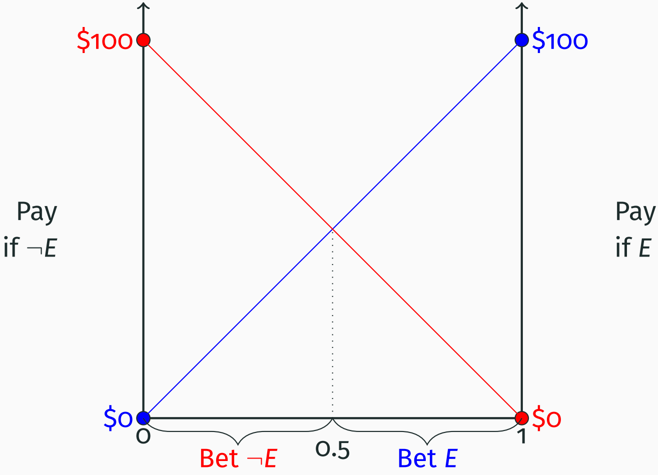
Assume (for now) risk-neutral EU

Theory: Savage (1971)



These two bets separate beliefs into two groups

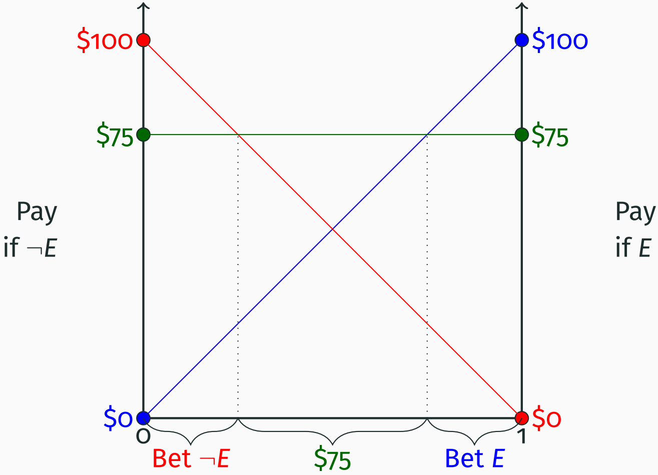
Theory: Savage (1971)



These two bets separate beliefs into two groups

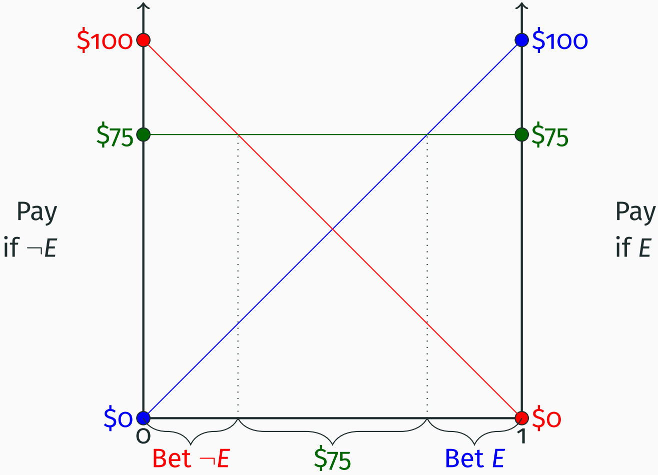
Revelation Principle: “Is $p \leq 0.5$ or is $p \geq 0.5$?”

Theory: Savage (1971)



We can get a finer elicitation by adding a constant bet!

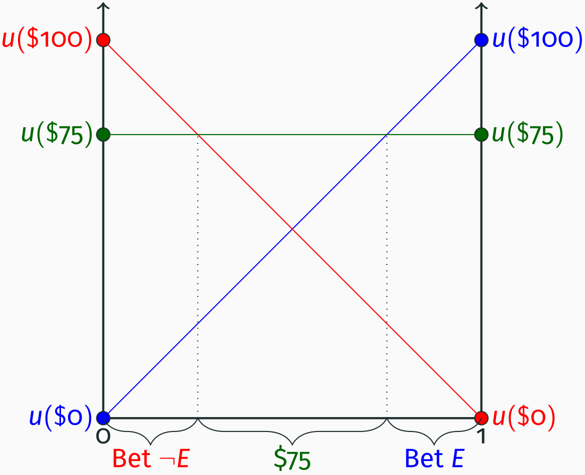
Theory: Savage (1971)



We can get a finer elicitation by adding a constant bet!

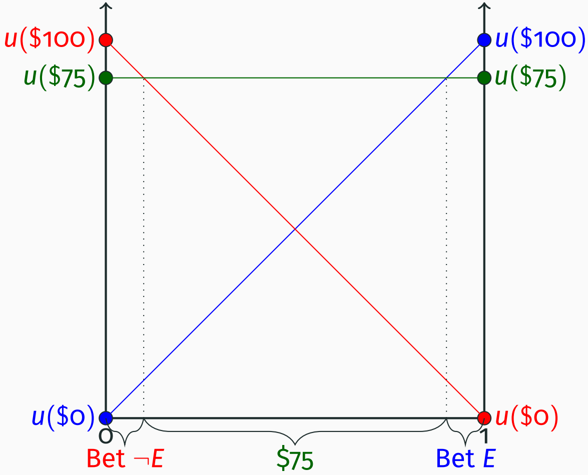
But what about risk aversion...?

Theory: Savage (1971)



Risk neutral

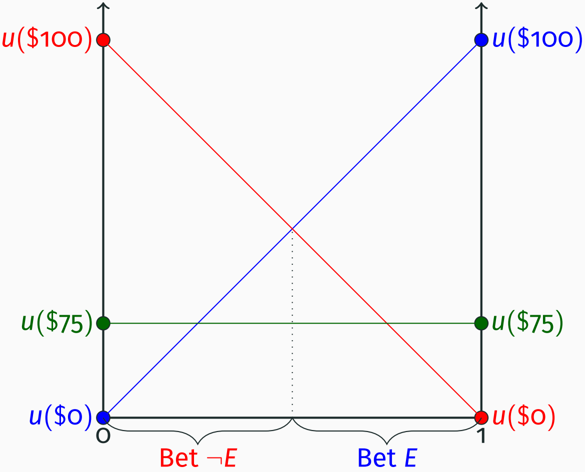
;



Risk averse

;

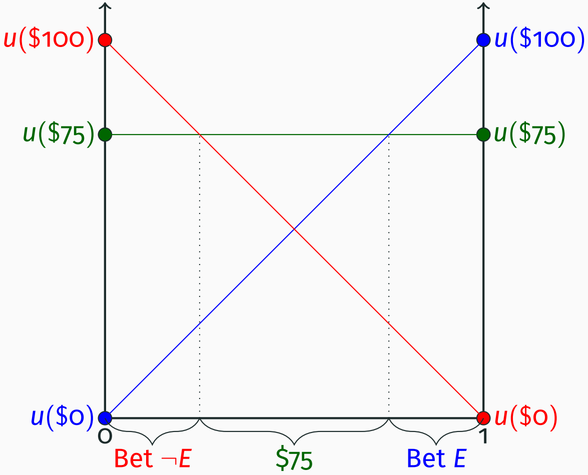
Theory: Savage (1971)



Risk seeking

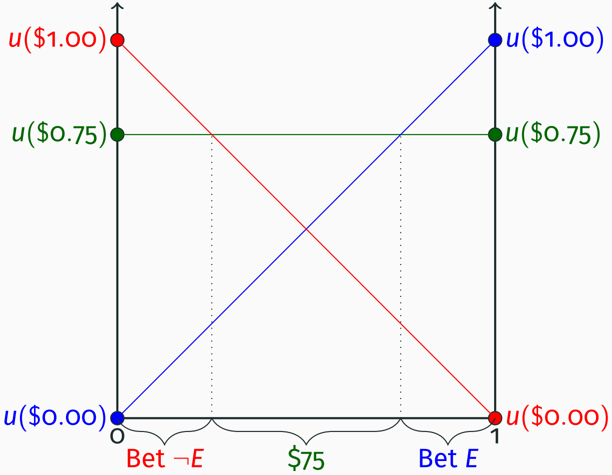
Risk preferences ⇒ lack of identification;

Theory: Savage (1971)



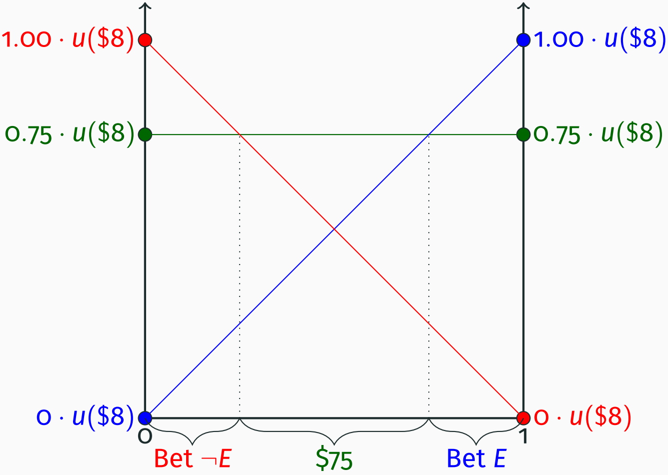
Savage (1971) offers 2 solutions...

Theory: Savage (1971)



Solution #1: make payments small ($1.00)

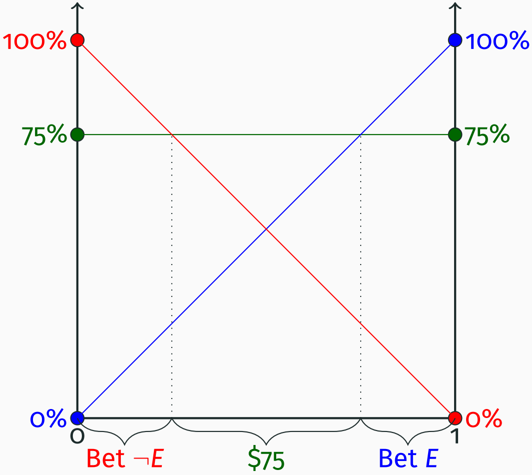
Theory: Savage (1971)



Solution #2: pay in probabilities

Payment = % chance of winning $8 (*e.g.*)

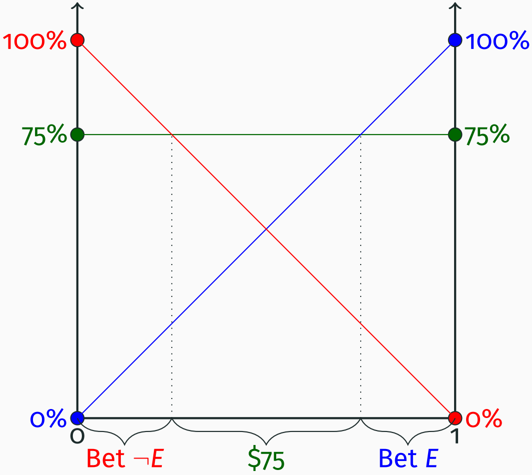
Theory: Savage (1971)



"Binarized" payments (Hossain & Okui 2013)

Savage (1971) → C. Smith (1961) → Savage (1954)

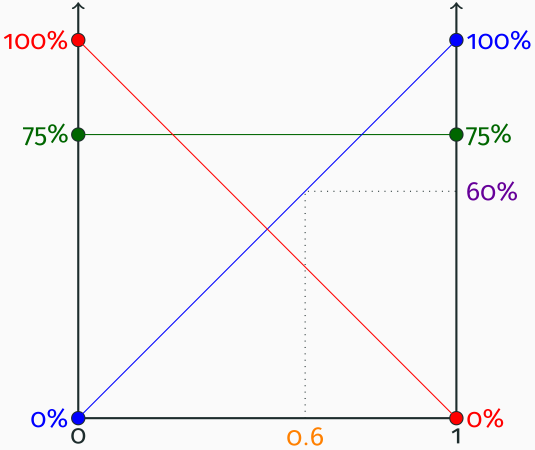
Theory: Savage (1971)



Solution #3: estimate risk prefs & back out *p*

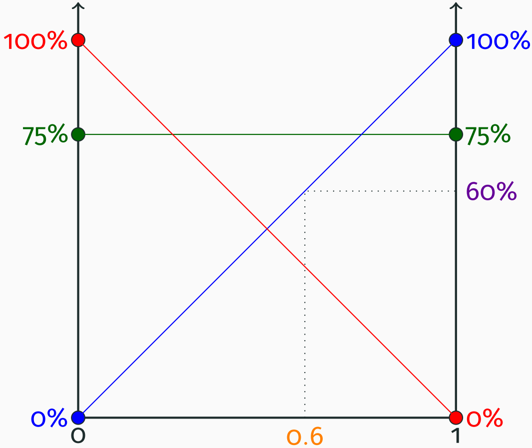
Offerman et al. (2009), Andersen et al. (2014), etc.

Theory: Savage (1971)



Still assuming linear preferences: $(0.6 \times 100\%) + (0.4 \times 0\%) = 60\%$

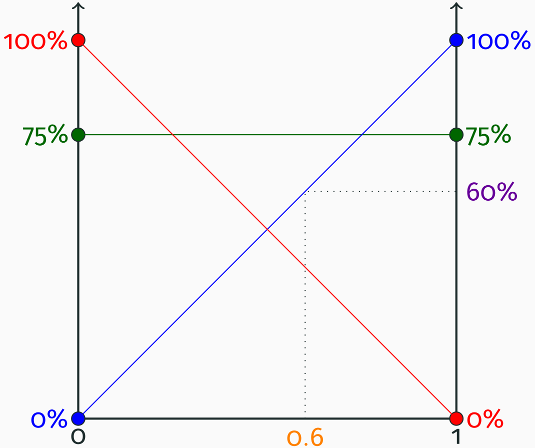
Theory: Savage (1971)



Still assuming linear preferences: $(0.6 \times 100\%) + (0.4 \times 0\%) = 60\%$

"Subjective-Objective Reduction" (aka Binary Reduction)

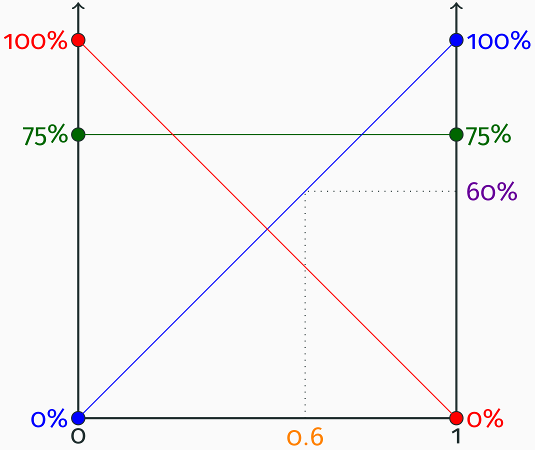
Theory: Savage (1971)



"Subjective-Objective Reduction"

Experimental evidence is pretty negative (Selten et al. 1999, *e.g.*)

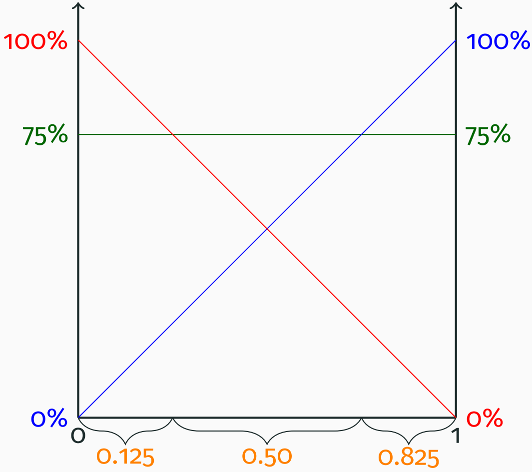
Theory: Savage (1971)



"Subjective-Objective Reduction"

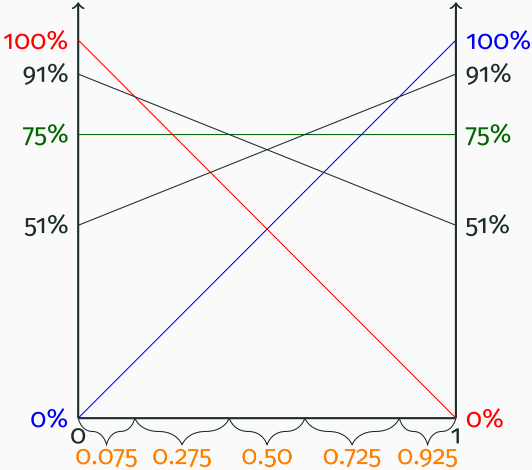
...except in the case of scoring rules (Hossain & Okui 2013, *e.g.*)

Theory: Savage (1971)



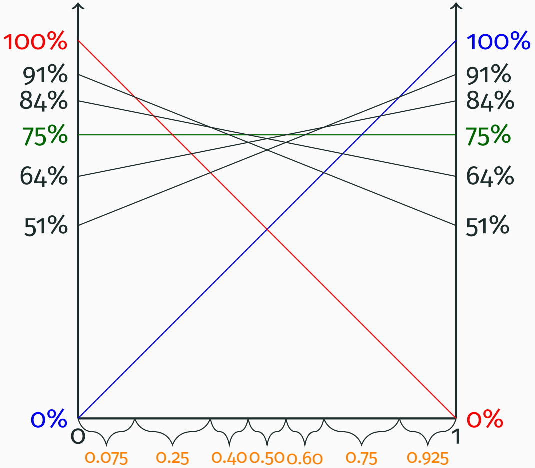
Now, let's add even more options to the menu...

Theory: Savage (1971)



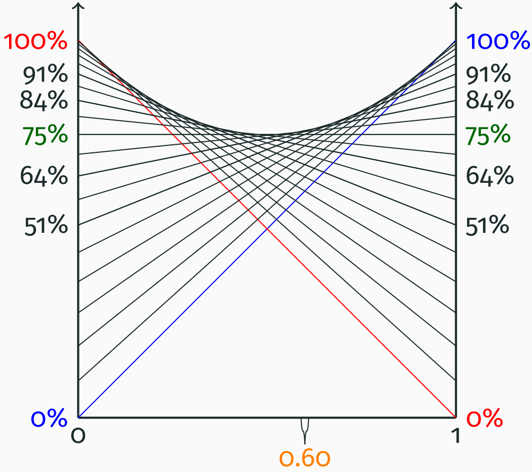
Now, let's add even more options to the menu...
5 categories

Theory: Savage (1971)



Now, let's add even more options to the menu...
7 categories

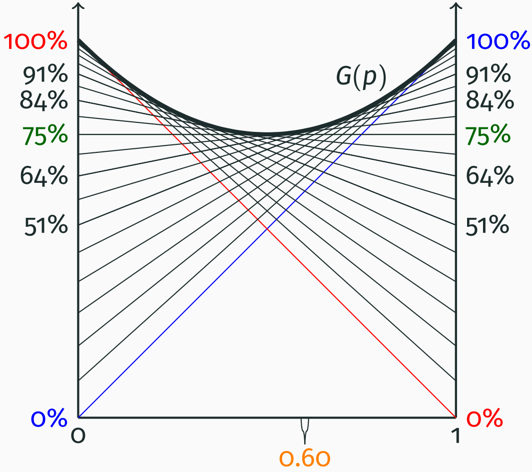
Theory: Savage (1971)



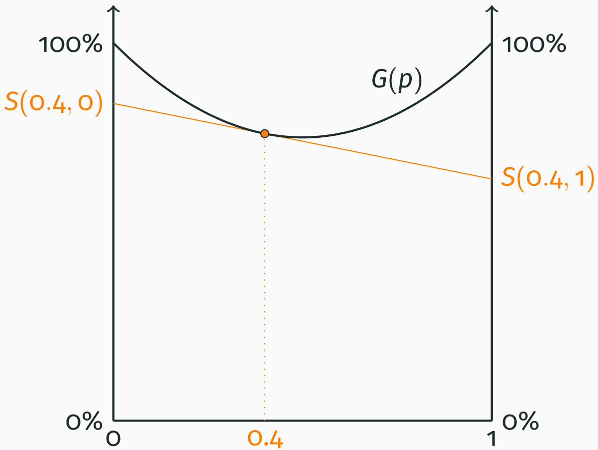
Now, let's add even more options to the menu...

↑ # bets → can elicit an exact p

Theory: Savage (1971)



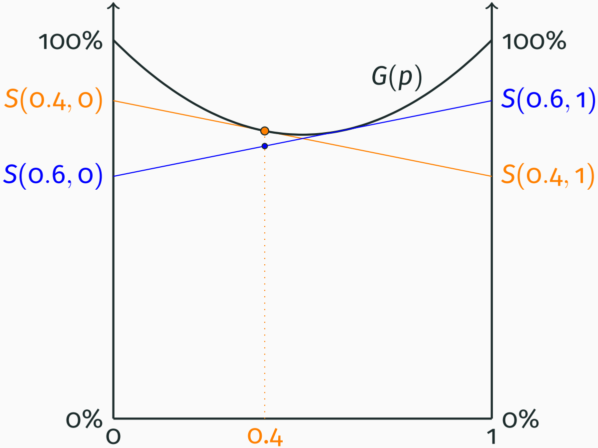
Convex upper envelope: $G(p)$
Each line is a tangent



Scoring Rule: Announce q.

If $\neg E$, pay $S(q, 0)$. If E, pay $S(q, 1)$.

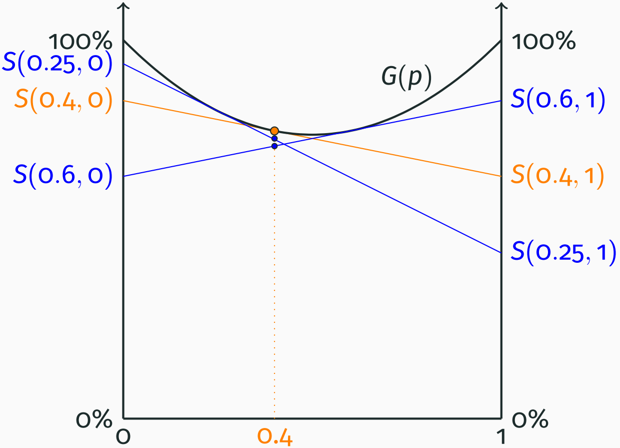
Theory: Savage (1971)



Scoring Rule: Announce q.

Announcing $q \neq p$ gives a lower $(1 - p) \cdot S(q, 0) + p \cdot S(q, 1)$

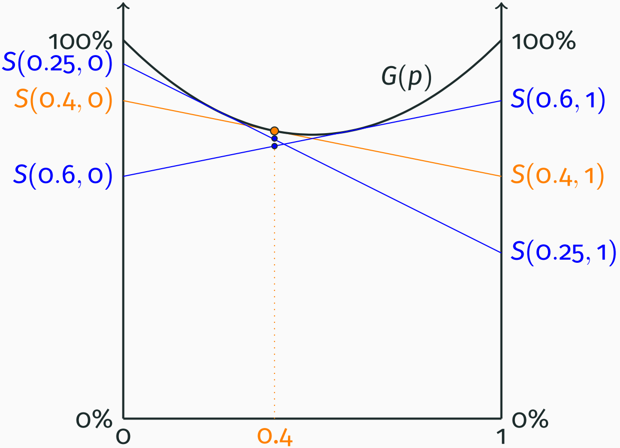
Theory: Savage (1971)



Scoring Rule: Announce q.

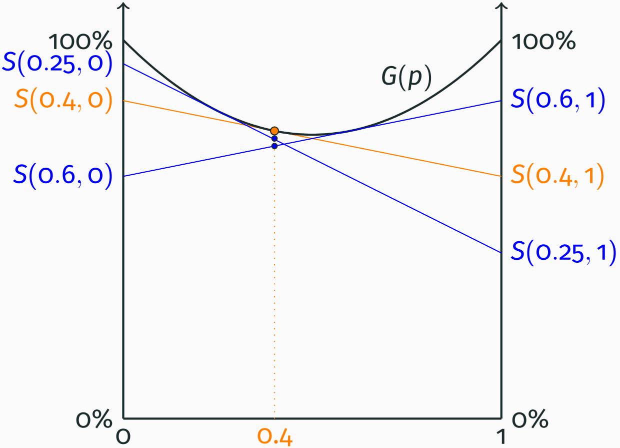
Announcing $q \neq p$ gives a lower $(1 - p) \cdot S(q, 0) + p \cdot S(q, 1)$

Theory: Savage (1971)



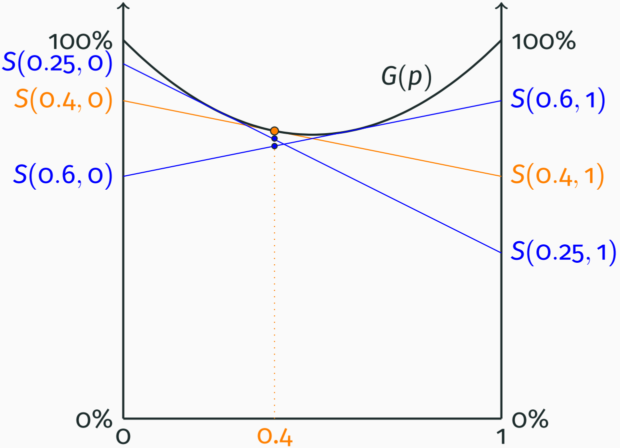
$$(1 - p) \cdot S(p, 0) + p \cdot S(p, 1)$$

Theory: Savage (1971)



Theorem (Savage/Schervish): A mechanism $S(p, X_E)$ is proper iff the resulting lines are the tangents of a convex function $G(p)$.

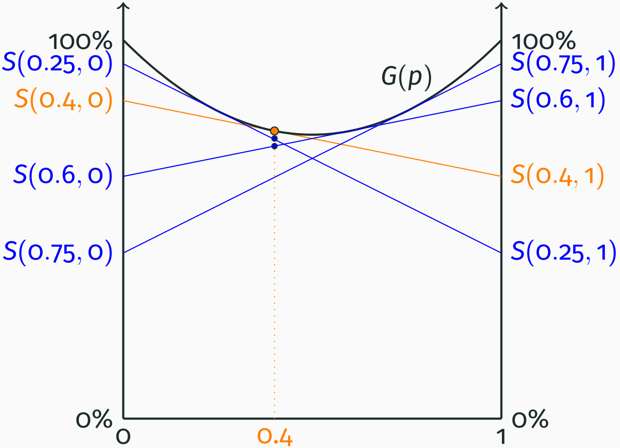
Theory: Savage (1971)



Any convex $G(p)$ will work.

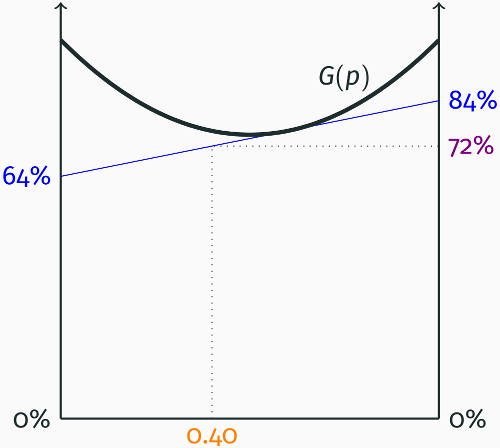
Binarized Quadratic scoring rule (BSR), logarithmic, spherical...

Theory: Savage (1971)



$$S(q,0) = (1 - (0-q)^2)$$
$$S(q,1) = (1 - (1-q)^2)$$

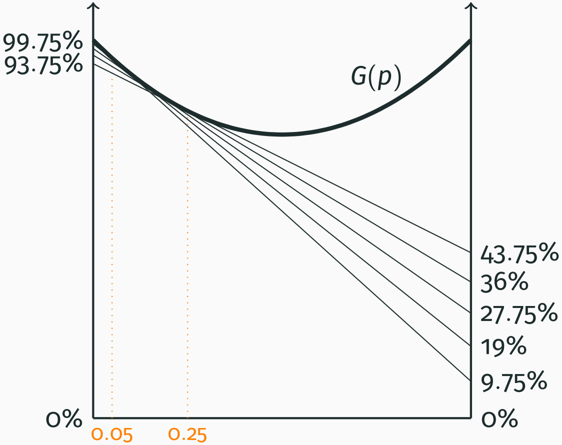
Issues With the Quadratic Scoring Rule



Concern #1: IC calculation requires S-O Reduction

$(0.4 \cdot 84\%) + (0.6 \cdot 64\%) = 72\%$

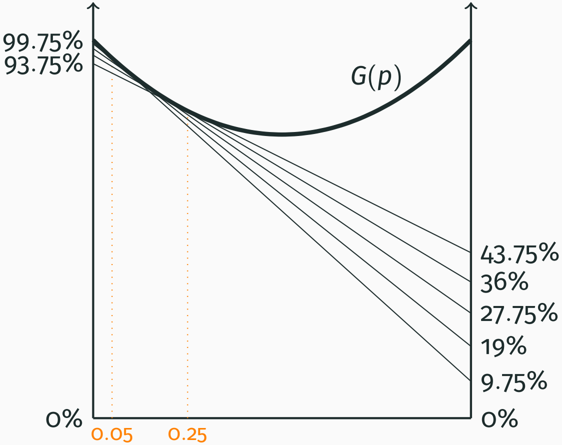
Issues With the Quadratic Scoring Rule



Concern #2: $S'(p, 0)$ vs $S'(p, 1)$

See Danz, Wilson & Vesterlund (2020), *e.g.*

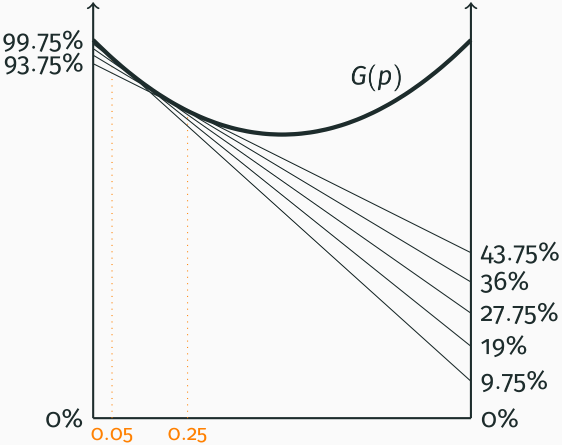
Issues With the Quadratic Scoring Rule



But see FOC: $pS'(p,1) + (1-p)S'(p,0) = 0$

$\Rightarrow p/(1-p) = -S'(p,0)/S'(p,1)$

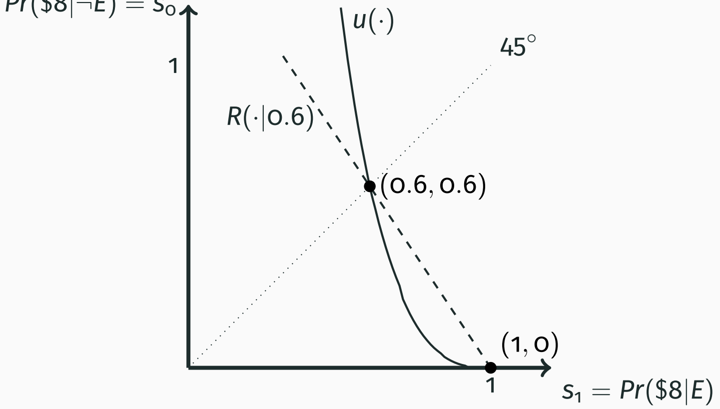
Issues With the Quadratic Scoring Rule



Relative slopes are pinned down by IC!

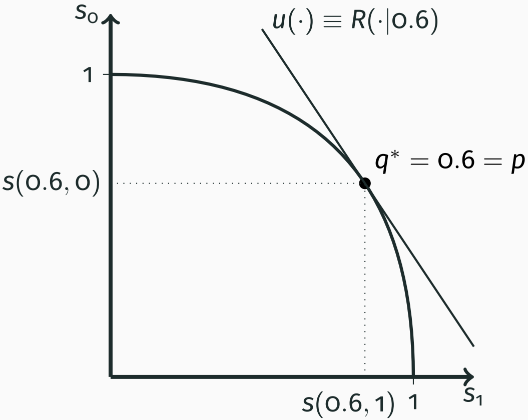
Corollary: For *any* IC scoring rule, $S'(p,0)/S'(p,1) = -p/(1-p)$.

An Alternative Visualization



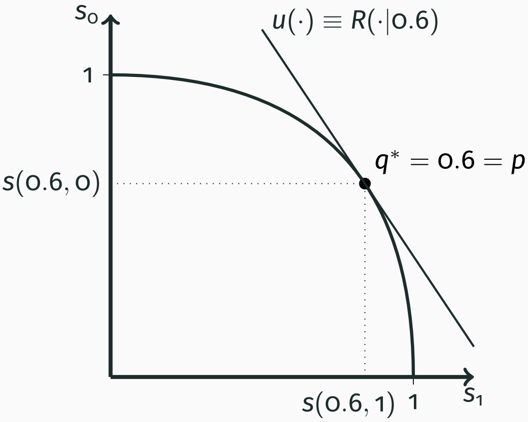
"Have a belief of 0.6": $u(1, 0) = u(0.6, 0.6)$
Define $R(s_1, s_0|p) = p \cdot s_1 + (1-p) \cdot s_0$. Linear level curves.
S-O Reduction: Have belief p and $u(s_1, s_0) = R(s_1, s_0|p)$

The BQSR



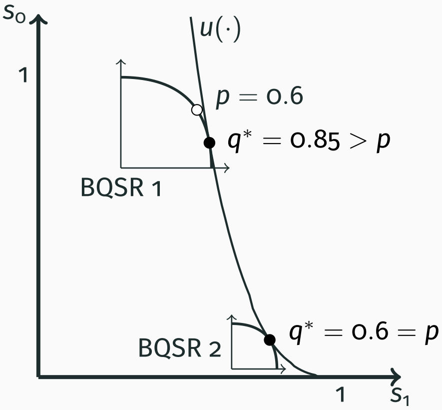
Binarized Quadratic Scoring Rule forms quarter-circle as you vary q
Maximizing point given $u(\cdot) \equiv R(\cdot|0.6)$ is $q^* = 0.6$

The BQSR



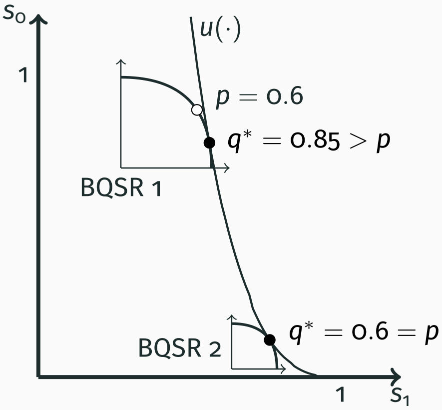
Any strictly concave shape corresponds to some proper scoring rule

Necessity of S-O Reduction



Know: If S-O Reduction then every scaled BQSR is IC
If $u(\cdot) \not\equiv R(\cdot|p)$ then $\exists$ scaled BQSR that's not IC.
Proposition: If every scaled BQSR is IC then $u(\dot{)} \equiv R(\cdot|p)$

Necessity of S-O Reduction



Know: If S-O Reduction then every scaled BQSR is IC

If $u(\cdot) \neq R(\cdot|p)$ then $\exists$ scaled BQSR that's not IC.

Proposition: If every scaled BQSR is IC then *S-O Reduction*

More Than One Event

- Suppose multiple events $E_1, E_2, \ldots, E_m$
- Want to elicit $p = (p_1, \ldots, p_m)$
- Let $X = i$ iff $\omega \in E_i$
- Announcement: $q = (q_1, \ldots, q_m)$

Quadratic Scoring Rule (scaled to $[0, 1]$):

$$S(q, i) = 1 - \frac{m}{m-1} \sum_{j=1}^{m} (\mathbb{1}_{\{X=j\}} - q_j)^2$$

Scaled BQSR:

$$S(q, i) = \beta_i - \alpha \sum_{j=1}^{m} (\mathbb{1}_{\{X=j\}} - q_j)^2$$

$$0 < \beta_j \leq 1 \;\; \forall j$$

$$0 < \alpha \leq \frac{m-1}{m} \min_j \beta_j$$

Other Scoring Rules

(These are not necessarily scaled to $[0, 1]$)

1. Spherical Scoring Rule (Roby 1964)

$$S(q, i) = \frac{q_i^2}{\sqrt{\sum_{j=1}^{m} q_j^2)}}$$

2. Generalized Spherical Scoring Rule ($\lambda > 1$)

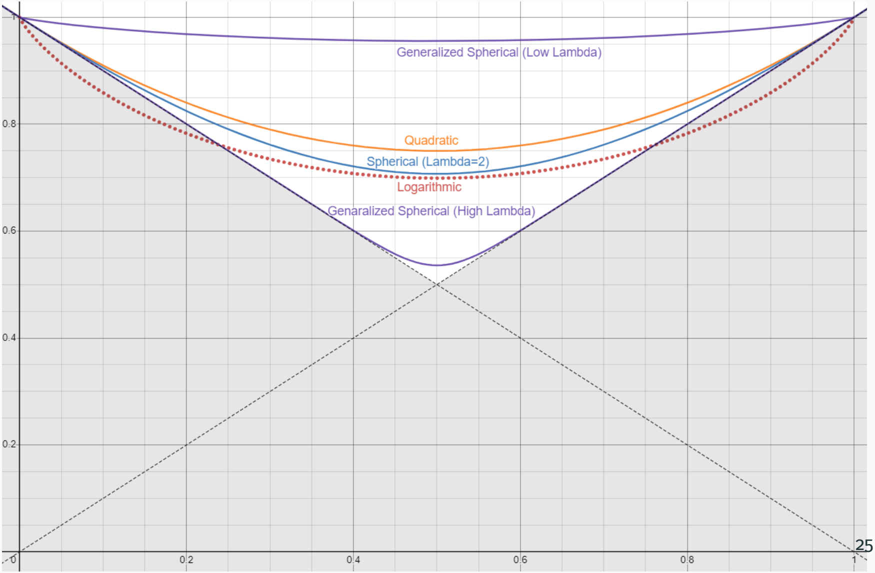
$$S(q, i) = \frac{q_i^\lambda}{(\sum_{j=1}^{m} q_j^\lambda)^{(\lambda-1)/\lambda}}$$

3. Logarithmic Scoring Rule

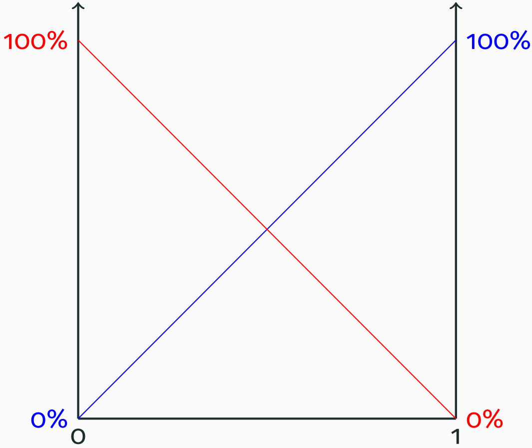
$$S(q, i) = \log q_i$$

(goes to $-\infty$, can't be scaled to $[0, 1]$)

Comparison of Scoring Rules



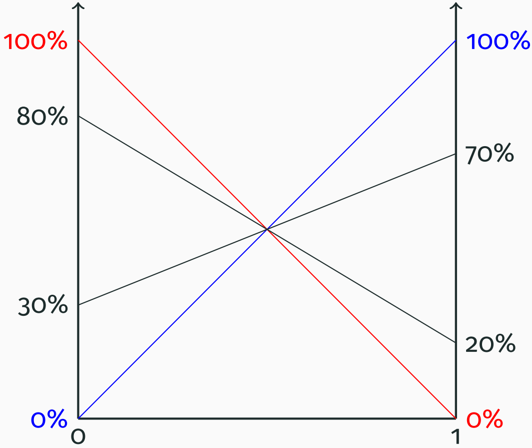
(Non-Proper) Linear Scoring Rule



$S(q, 0) = 1 - q$ $\qquad$ $S(q, 1) = q$

Same extremes as QSR

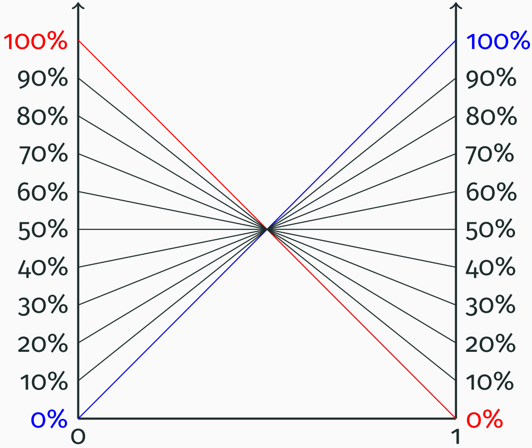
(Non-Proper) Linear Scoring Rule



$S(q, 0) = 1 - q$ $S(q, 1) = q$

But now symmetric slopes

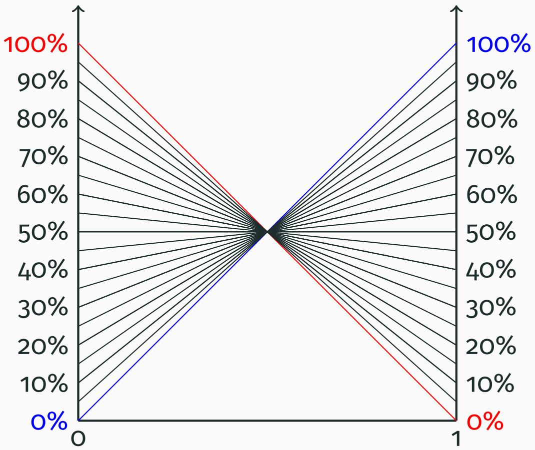
(Non-Proper) Linear Scoring Rule



$S(q, 0) = 1 - q$ $\qquad$ $S(q, 1) = q$

But now symmetric slopes

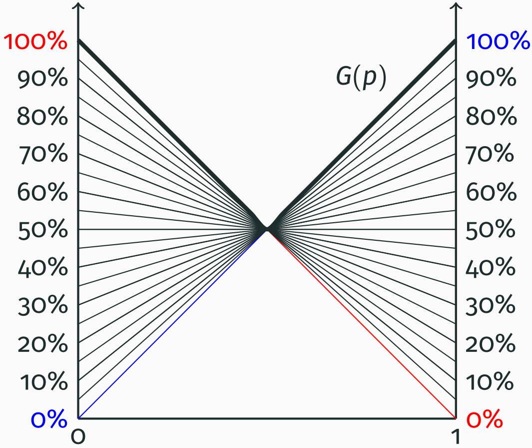
(Non-Proper) Linear Scoring Rule



$S(q, 0) = 1 - q$ $\qquad$ $S(q, 1) = q$

But now symmetric slopes

(Non-Proper) Linear Scoring Rule



Convex upper envelope: $G(p)$

$q^* = 0$ if $p < 50$, $q^* = 100$ if $p > 50$

Characterizations of the QSR

Selten (1998, *ExpEcon* v.1)

- **Symmetry**: $S(q, i) = S(\pi(q), \pi(i))$ for any permutation π
- **Elongation Invariance:** $S((q_1, \ldots, q_n), i) = S((q_1, \ldots, q_n, 0), i)$ (adding a null event)
- **Neutrality:** $G(q|p) = G(p|q)$
- **Properness:** S is proper

Theorem: A scoring rule satisfies these 4 axioms iff it is a scaled QSR

Characterizations of the QSR

- Suppose we impose a grid $\mathcal{G} = \{0, \frac{1}{k}, \frac{2}{k}, \ldots, \frac{k-1}{k}, 1\}$
- Require each $q_i \in \mathcal{G}$
- **Midpoint Property:** Optimal announcement is $q_i^* = \frac{r}{k}$ if and only if $p_i \in [\frac{r}{k} - \frac{1}{2k}, \frac{r}{k} + \frac{1}{2k}]$
  - Ensures that the announced point is the closest grid point to the true belief.

Theorem: The Scaled QSRs are the only proper scoring rules with the midpoint property

Characterizations of the QSR

- We want to maximize the incentive not to deviate
- Local incentive not to deviate at $q = p$

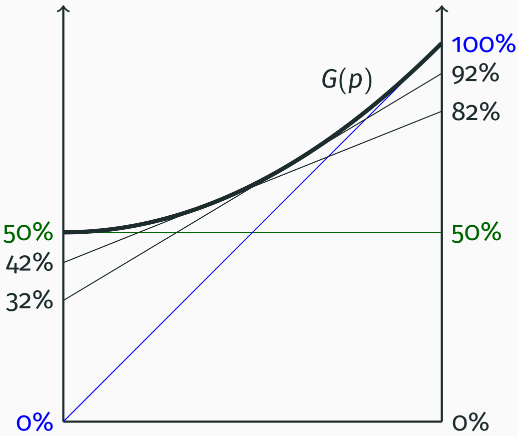
$$G''(q = p|p) = G''(p)$$

- BQSR has $G'' \equiv 2$
- Any binarized rule must have $G'(0) \geq -1$, $G'(1) \leq 1$
  - All lines in the graph must have slope in $[-1, 1]$
- Thus, $\int_0^1 G''(p)dp = G'(1) - G'(0) \leq 2$
- Any other scoring rule has $G''(p) < 2$ at some p

Theorem: The (unscaled) BQSR maximizes $\min_p G''(p)$

Related: Schlag, Tremewan & van der Weele (2015)

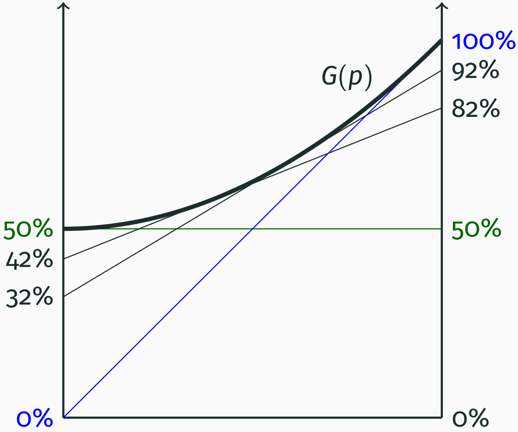
A Different Scoring Rule



A new IC scoring rule

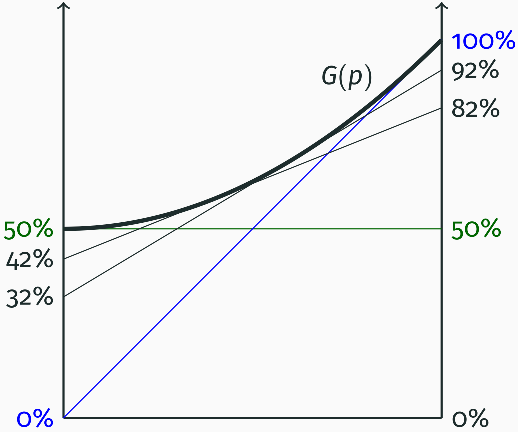
$G(p) = \frac{1}{2}(1 + q^2)$

A Different Scoring Rule



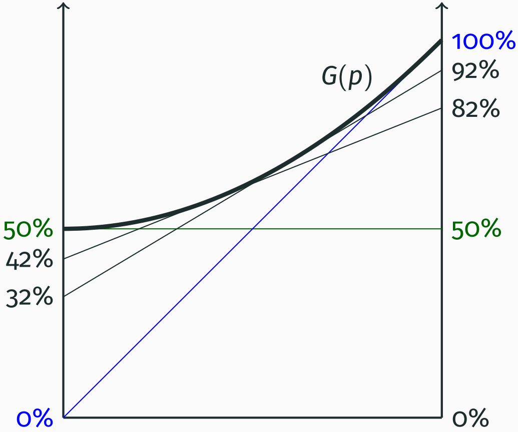
$$S(q,0) = \frac{1}{2}(1 - q^2)$$
$$S(q,1) = \frac{1}{2}(1 - (1-q)^2) + \frac{1}{2}$$

A Different Scoring Rule



$$S(q,0) = \frac{1}{2}(1 - q^2)$$
$$S(q,1) = \frac{1}{2}(1 - (1-q)^2) + \frac{1}{2}$$

A Different Scoring Rule



Magic Trick: I'll show *this* scoring rule can be IC *without* relying on S-O Reduction

Breaking Apart Reduction

Consider the S-O-Reduced $Pr(\$8)$:

$$p \cdot (\underbrace{\frac{1}{2}(1-(1-q)^2)+\frac{1}{2}}_{S(q,1)}) + (1-p) \cdot \underbrace{\frac{1}{2}(1-q^2)}_{S(q,0)}$$

$$= q \cdot p + (1-q)\left(\frac{1}{2}q + \frac{1}{2}1\right)$$

Breaking Apart Reduction

Consider the S-O-Reduced $Pr(\$8)$:

$$p \cdot (\underbrace{\frac{1}{2}(1-(1-q)^2)+\frac{1}{2}}_{S(q,1)}) + (1-p) \cdot \underbrace{\frac{1}{2}(1-q^2)}_{S(q,0)}$$

$$= q \cdot p + (1-q)\left(\frac{1}{2}q + \frac{1}{2}1\right)$$

q : get a \$8 bet on E

$(1-q)$: get a lottery that pays \$8 w/ prob $\left(\frac{1}{2}q + \frac{1}{2}1\right)$

Breaking Apart Reduction

Consider the S-O-Reduced $Pr(\$8)$:

$$p \cdot (\underbrace{\frac{1}{2}(1-(1-q)^2)+\frac{1}{2}}_{S(q,1)}) + (1-p) \cdot \underbrace{\frac{1}{2}(1-q^2)}_{S(q,0)}$$

$$= q \cdot p + (1-q)\left(\frac{1}{2}q + \frac{1}{2}1\right)$$

q : get a \$8 bet on E

$(1-q)$: get a lottery that pays \$8 w/ prob $\left(\frac{1}{2}q + \frac{1}{2}1\right)$

Adding a second objective randomizing device

Breaking Apart Reduction

$$q \cdot p + (1 - q)\frac{q+1}{2}$$

Imagine 100 rows. Announce $q \in [0, 100]$. Payment:

$$q \left\{ \begin{array}{l} \$8 \text{ if } E \\ \$8 \text{ if } E \\ \vdots \\ \$8 \text{ if } E \end{array} \right\} = q \cdot p\%$$

$$(1-q) \left\{ \begin{array}{c} \$8 \text{ w/ prob } q + 1\% \\ \$8 \text{ w/ prob } q + 2\% \\ \vdots \\ \$8 \text{ w/ prob } 99\% \\ \$8 \text{ w/ prob } 100\% \end{array} \right\} = (1-q) \cdot \underbrace{\left(\frac{1}{2}q + \frac{1}{2}1\right)}_{\text{Avg. prob. from } q \text{ to } 1} \%$$

Breaking Apart Reduction: Multiple Price List

Row#	Option A	OR	Option B
1	\$8 if E	or	\$8 w/ prob 1%
2	\$8 if E	or	\$8 w/ prob 2%
⋮	⋮	⋮	⋮
q	\$8 if E	or	\$8 w/ prob q%
$q+1$	\$8 if E	or	\$8 w/ prob $q+1$%
⋮	⋮	⋮	⋮
99	\$8 if E	or	\$8 w/ prob 99%
100	\$8 if E	or	\$8 w/ prob 100%

Equivalently: Choose Option A or Option B

Choice of q determines your choices

Breaking Apart Reduction: Multiple Price List

Row#	Option A	OR	Option B
1	\$8 if E	or	\$8 w/ prob 1%
2	\$8 if E	or	\$8 w/ prob 2%
⋮	⋮	⋮	⋮
q	\$8 if E	or	\$8 w/ prob q%
$q+1$	\$8 if E	or	\$8 w/ prob $q+1$%
⋮	⋮	⋮	⋮
99	\$8 if E	or	\$8 w/ prob 99%
100	\$8 if E	or	\$8 w/ prob 100%

"Multiple Price List" (MPL) version of BDM for probabilities
Holt & Smith (2016), others

Breaking Apart Reduction: Multiple Price List

Row#	**Option A**	OR	**Option B**
1	\$8 if E	or	\$8 w/ prob 1%
2	\$8 if E	or	\$8 w/ prob 2%
⋮	⋮	⋮	⋮
q	\$8 if E	or	\$8 w/ prob q%
$q+1$	\$8 if E	or	\$8 w/ prob $q+1$%
⋮	⋮	⋮	⋮
99	\$8 if E	or	\$8 w/ prob 99%
100	\$8 if E	or	\$8 w/ prob 100%

One row randomly selected for payment

If you lie, you get the less-preferred option on some rows

Breaking Apart Reduction: Multiple Price List

Row#	Option A	OR	Option B
1	\$8 if E	or	\$8 w/ prob 1%
2	\$8 if E	or	\$8 w/ prob 2%
⋮	⋮	⋮	⋮
q	\$8 if E	or	\$8 w/ prob q%
$q+1$	\$8 if E	or	\$8 w/ prob $q+1$%
⋮	⋮	⋮	⋮
99	\$8 if E	or	\$8 w/ prob 99%
100	\$8 if E	or	\$8 w/ prob 100%

One row randomly selected for payment

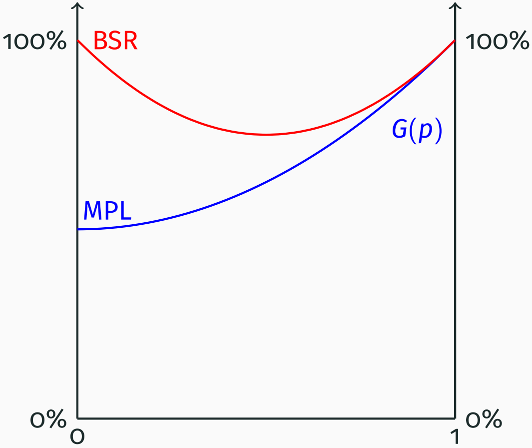
I.C. as long as subject respects statewise dominance

Breaking Apart Reduction: Multiple Price List

Row#	Option A	OR	Option B
1	\$8 if E	or	\$8 w/ prob 1%
2	\$8 if E	or	\$8 w/ prob 2%
⋮	⋮	⋮	⋮
q	\$8 if E	or	\$8 w/ prob q%
$q+1$	\$8 if E	or	\$8 w/ prob $q+1$%
⋮	⋮	⋮	⋮
99	\$8 if E	or	\$8 w/ prob 99%
100	\$8 if E	or	\$8 w/ prob 100%

Summary: Took a scoring rule, converted it into an MPL
Now IC does *not* require S-O Reduction!

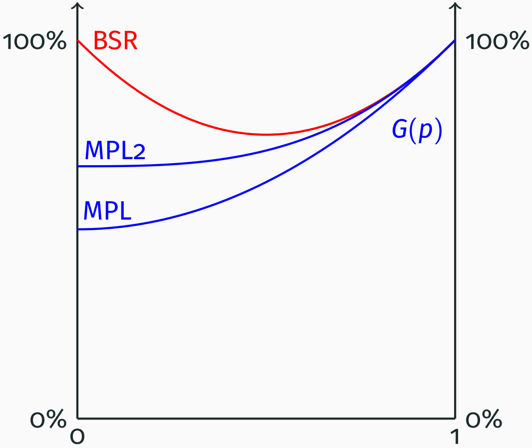
What Can Be Listified?



Proposition: $G(p)$ can be made into an MPL if and only if

1. $G'(0) = 0$ 2. $G'(1) = 1$ 3. $G(1) = 1$

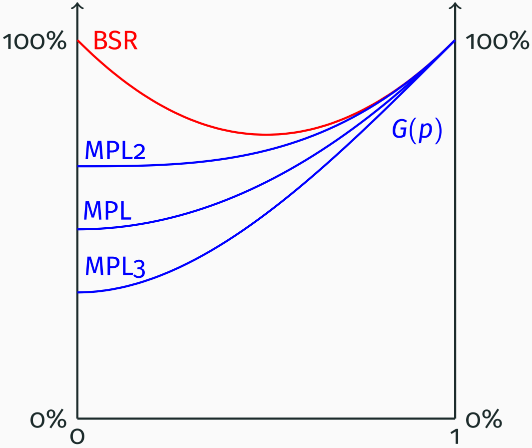
What Can Be Listified?



Proposition: $G(p)$ can be made into an MPL if and only if

1. $G'(0) = 0$ 2. $G'(1) = 1$ 3. $G(1) = 1$

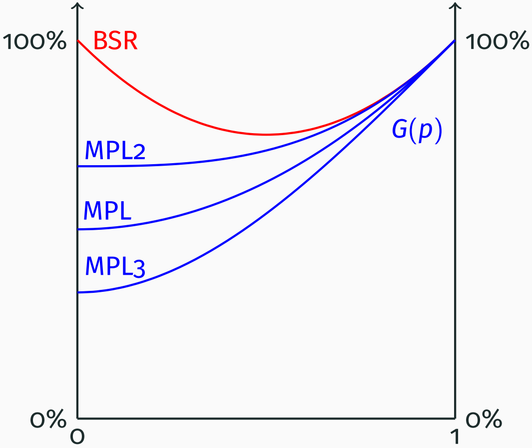
What Can Be Listified?



Proposition: $G(p)$ can be made into an MPL if and only if

1. $G'(0) = 0$
2. $G'(1) = 1$
3. $G(1) = 1$

What Can Be Listified?



What's the difference across MPLs?
Varying probability of rows being chosen

Superiority of MPLs

We can argue that the MPLs are superior to the BQSRs:

Theorem:

All Scaled BQSRs are I.C.

$$\Updownarrow$$

Subjective-Objective Reduction

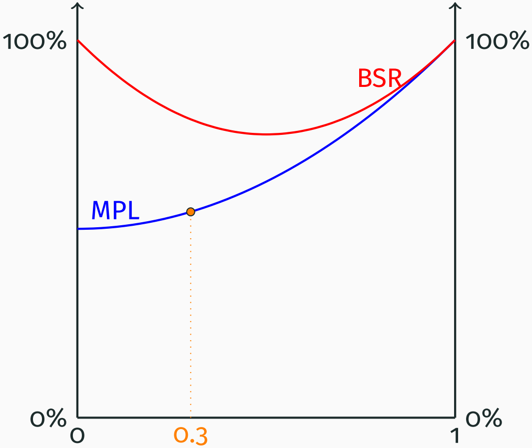
$$\Downarrow$$

Statewise Dominance

$$\Downarrow$$

MPL is I.C. (regardless of dist'n on rows)

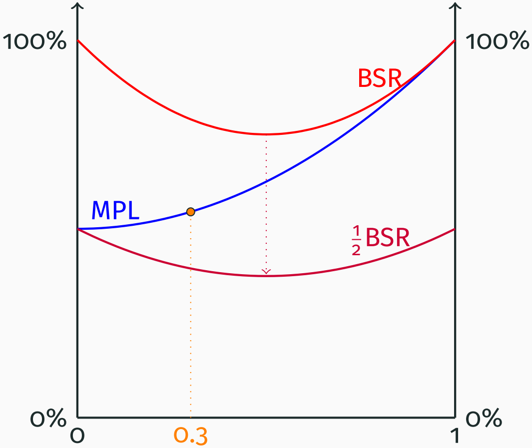
Equalizing Incentives



How to equalize incentives across scoring rules?

e.g. suppose we know $p = 0.3$

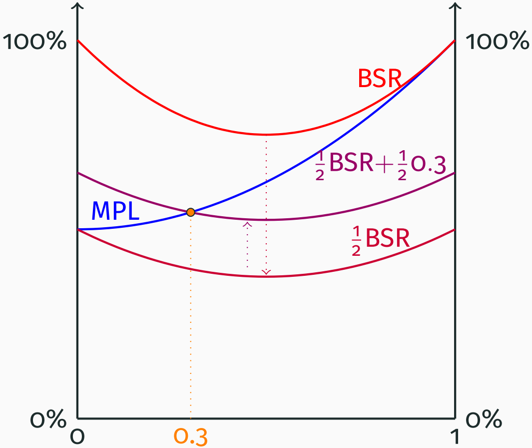
Equalizing Incentives



How to equalize incentives across scoring rules?

e.g. suppose we know $p = 0.3$

Equalizing Incentives



How to equalize incentives across scoring rules?

Heads: use BSR. Tails: get $8 w/ prob 0.3.

Equalizing Incentives

- Let X be r.v. representing E
  - $E \Rightarrow X = 1$
  - $\neg E \Rightarrow X = 0$
- MPL:

$$S(p, x) = \frac{1}{2}(1 - (x - p)^2) + \frac{1}{2}x$$

- Suppose researcher's best guess of p is p_0
- Adjusted BSR:

$$S(p, x) = \frac{1}{2}(1 - (x - p)^2) + \frac{1}{2}p_0$$

Equalizing Incentives

- Let X be r.v. representing E
  - $E \Rightarrow X = 1$
  - $\neg E \Rightarrow X = 0$
- MPL:

$$S(p, x) = \frac{1}{2}(1 - (x - p)^2) + \frac{1}{2}x$$

- Suppose researcher's best guess of p is p_0
- Adjusted BSR:

$$S(p, x) = \frac{1}{2}(1 - (x - p)^2) + \frac{1}{2}p_0$$

Other Statistics of a Distribution

- Consider again general r.v. X
  - BSR: $S(p, x) = \left(1 - (x - p)^2\right)$
- Can we elicit a statistic of p? Ex: mean, median, mode, ...
- Could elicit $Pr(X = x)$ for every possible x... but that's a lot!
- The (single-report) BSR elicits the subject's **mean** for X
  - BSR: $S(m, x) = \left(1 - (x - m)^2\right)$
  - Still paying in probabilities
  - Still requiring S-O Reduction:

$$\max_{m} \sum_{x} Pr(X = x)(1 - (x - m)^2)$$

- Can we elicit the mean using an MPL?

MPL for The Mean of X

Row#	Option A	OR	Option B
1	X% chance of \$8	or	1% chance of \$8
2	X% chance of \$8	or	2% chance of \$8
⋮	⋮	⋮	⋮
m	X% chance of \$8	or	m% chance of \$8
m+1	X% chance of \$8	or	m+1% chance of \$8
⋮	⋮	⋮	⋮
99	X% chance of \$8	or	99% chance of \$8
100	X% chance of \$8	or	100% chance of \$8

Identical to two-state list: Option A is (\$8 if E)
but, now requires linearity: "X% chance" $\sim$ "$E[X]$% chance"

MPL for The Mean of X

Row#	Option A	OR	Option B
1	X% chance of \$8	or	1% chance of \$8
2	X% chance of \$8	or	2% chance of \$8
⋮	⋮	⋮	⋮
m	X% chance of \$8	or	m% chance of \$8
m+1	X% chance of \$8	or	m+1% chance of \$8
⋮	⋮	⋮	⋮
99	X% chance of \$8	or	99% chance of \$8
100	X% chance of \$8	or	100% chance of \$8

Now requires linearity: "X% chance" $\sim$ "$E[X]$% chance"
but, given that, IC only requires statewise dominance

Equalizing Incentives with Mean Elicitation

- Researcher's best guess: mean is μ_0, variance is σ_0^2
- (Recall: $E[X^2] = \mu_0^2 + \sigma_0^2$)
- BSR:

$$S(p, x) = \quad (1 - (x - m)^2)$$

- MPL:

$$S(p, x) = \frac{1}{2}(1 - (x - m)^2) + \frac{1}{2}x^2$$

Equalizing Incentives with Mean Elicitation

- Researcher's best guess: mean is μ_0, variance is σ_0^2
- (Recall: $E[X^2] = \mu_0^2 + \sigma_0^2$)
- BSR:

$$S(p,x) = \frac{1}{2}(1-(x-m)^2) + \frac{1}{2}(\mu_0^2 + \sigma_0^2)$$

- MPL:

$$S(p,x) = \frac{1}{2}(1-(x-m)^2) + \frac{1}{2}x^2$$

Eliciting the Median

- BSR elicits the mean... can we elicit the median?
- **Linear** scoring rule elicits the median!
- LSR:

$$S(m, x) = (1 - |x - m|)$$

- Can **this** be listified?

MPL for The Median of X

Row#	Option A	OR	Option B
1	$8 if $X \geq 1$	or	50% chance of $8
2	$8 if $X \geq 2$	or	50% chance of $8
⋮	⋮	⋮	⋮
m	$8 if $X \geq m$	or	50% chance of $8
m+1	$8 if $X \geq m+1$	or	50% chance of $8
⋮	⋮	⋮	⋮
99	$8 if $X \geq 99$	or	50% chance of $8
100	$8 if $X \geq 100$	or	50% chance of $8

Does *NOT* require linearity

Easily altered to elicit any quantile

Equalizing Incentives with Median Elicitation

- Suppose researcher's best guess of the median is $\mu_{0.5}$
- BSR:

$$S(p,x) = \quad (1 - |x - m|)$$

- MPL:

$$S(p,x) = \frac{1}{2}(1 - |x - m|) + \frac{1}{2}x$$

Equalizing Incentives with Median Elicitation

- Suppose researcher's best guess of the median is $\mu_{0.5}$
- BSR:

$$S(p,x) = \frac{1}{2}(1 - |x - m|) + \frac{1}{2}\mu_{0.5}$$

- MPL:

$$S(p,x) = \frac{1}{2}(1 - |x - m|) + \frac{1}{2}x$$

Eliciting the Mode

- Eliciting the mode is simple & stark:

$$S(m, x) = \mathbb{1}_{x=m}$$

- Generally: elicit most-likely interval of length d
  - Announce any $[\underline{m}, \overline{m}]$ s.t. $\overline{m} - \underline{m} = d$

$$S([\underline{m}, \overline{m}], x) = \mathbb{1}_{x \in [\underline{m}, \overline{m}]}$$

  - Use this if X has many values, since $Pr(x = m) \approx 0 \ \ \forall m$

Scoring Rules for Quantiles

- We saw MPLs can be used to elicit quantiles
- Scoring rule for eliciting α quantile (Cervera & Muñoz 1996):

$$S(m, x) = \alpha m - (m - x)\mathbb{1}_{x \leq m}$$

- Median is $\alpha = 1/2$
- Proof: True distribution is $p(x)$

$$\int S(m, x)p(x)dx = \alpha m - \int_0^m (m - x)p(x)dx$$

$$FOC : \alpha - (m - m)p(m) - \int_0^m 1\, p(x)dx = 0$$

- Announce m such that $\int_0^m p(x)dx = \alpha$

Eliciting Confidence Intervals

- We want to elicit the 95% confidence interval
- Separately elicit 2.5% quantile and 97.5% quantile
- Pay one elicitation randomly

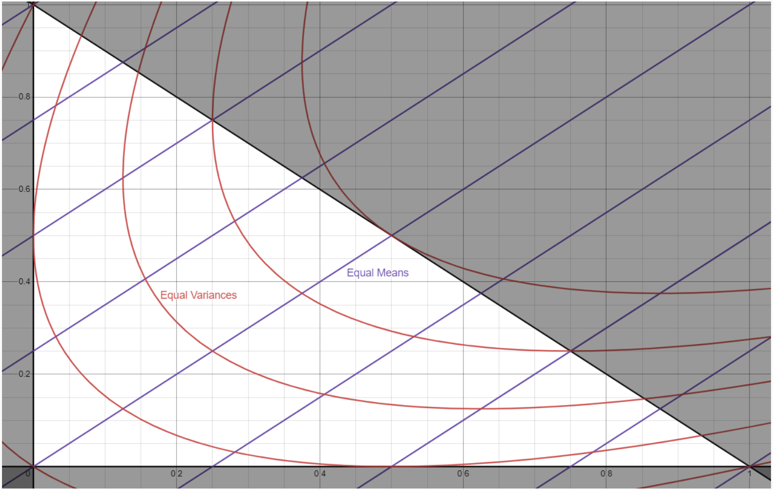
The Lambert Characterization

Lambert, Pennock & Shoham (2008)

- In general, a *statistic* is a mapping $\Gamma : \Delta(\Omega) \to \mathbb{R}$
- Examples: mean, median, mode, variance, kurtosis...
- What statistics can be elicited?

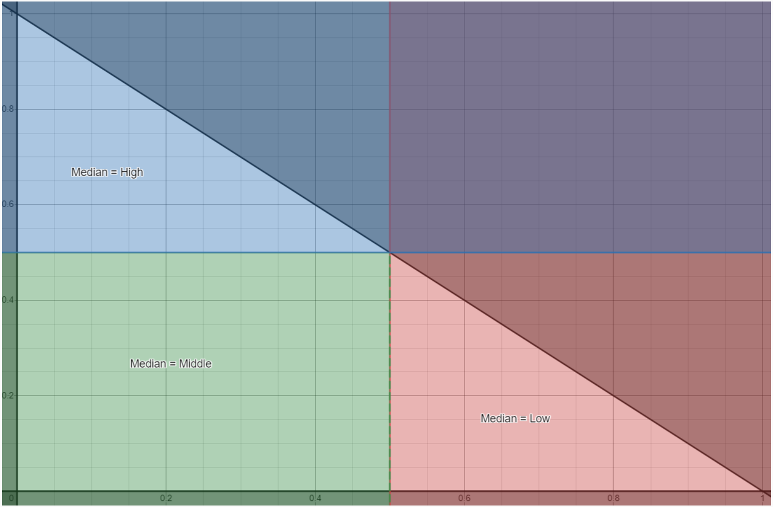
Theorem: A statistic Γ can be elicited via a strictly proper scoring rule if and only if $\Gamma^{-1}(r)$ is a convex set of distributions for every possible statistic value r

The Lambert Characterization



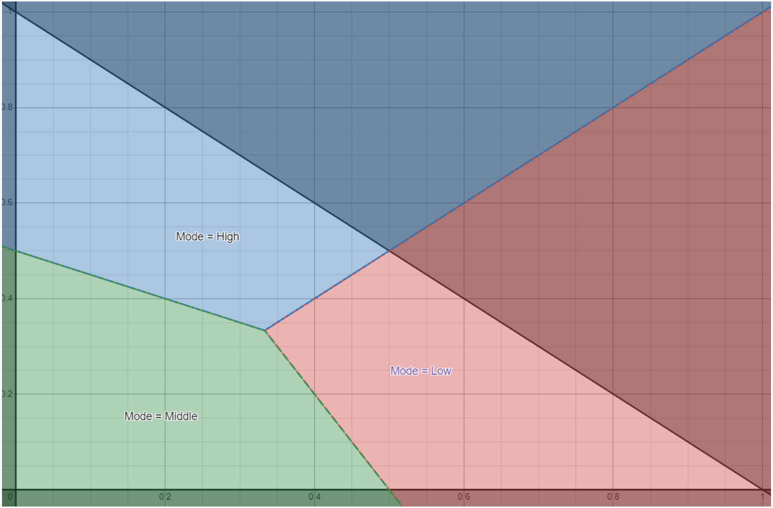
Mean: yes. Variance: no!

The Lambert Characterization



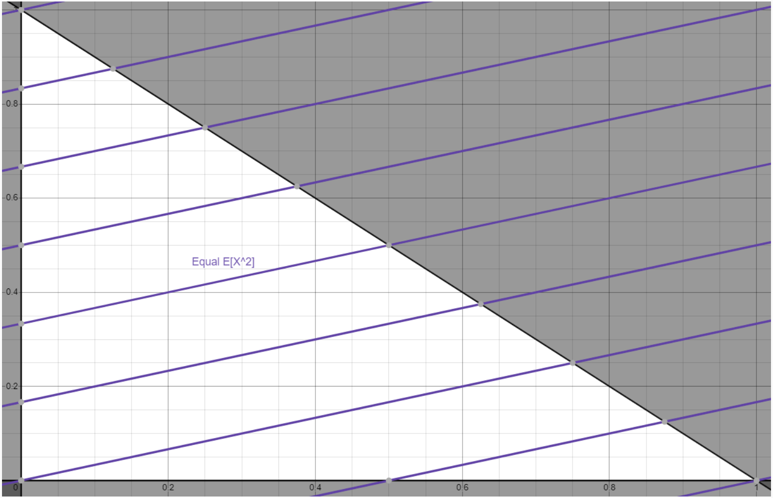
Median: yes!

The Lambert Characterization



Mode: yes!

The Lambert Characterization



$E[X^2]$: yes! (Why do we care?? Next slide...)

The Lambert Characterization

- We can't elicit $Var_p(X)$ with 1 report
- But we can elicit $E_p(X)$ and $E_p(X^2)$
  - $Var_p(X) = E_p(X^2) - E_p(X)^2$
- Or, suppose we observe two draws X_1 and X_2 from same dist'n
- Then $X_1 - X_2$ is a new r.v.
- We can elicit $E_p((X_1 - X_2)^2)$
  - $Var_p(X) = E_p((X_1 - X_2)^2)$ (check this)

Survey of Experimental Results

Schotter & Trevino (2014)
Does IC matter?

1. Nelson & Bessler (1989)
    - Only use risk-neutral subjects
    - Compare BSR to non-IC Linear SR
    - Early periods: same. Later: differences
2. Palfrey & Wang (2009)
    - QSR vs LogSR vs LinearSR in games
    - Beliefs elicited via IC mechanism are better forecasts

Survey of Experimental Results

Schotter & Trevino (2014)
Risk aversion and the standard QSR:

1. Armantier & Treich (2013)
    - Theoretical predictions for what should happen under risk aversion
    - Observe predicted “flatness” in reports
    - No incentives increases variance of reports
2. Offerman & Sonnemans (2004)
    - QSR performs same as flat fee
3. Hossain & Okui (2013)
    - BQSR outperforms QSR

Survey of Experimental Results

Schotter & Trevino (2014)
Do people best-reply to stated beliefs in games?

1. Nyarko & Schotter (2002): yes, BR is most likely
2. Rey-Biel (2009) 3×3 games: yes, 69.4%
3. Blanco et al. (2011) seq. PD: yes
4. Hyndman et al. (2013): yes, even days later
5. Danz et al. (2012) 3×3: yes
6. Ivanov (2011): yes
7. Manski & Neri (2013): yes
8. Costa-Gomes & Weizsacker (2008)
    - 14 3×3 games
    - Trt: games-then-elicitations vs. both together
    - Can we back out beliefs from actions and match stated beliefs?
    - Result: NO

Survey of Experimental Results

Schotter & Trevino (2014)
Does elicitation change subsequent behavior?

1. Nyarko & Schotter (2002): no
2. Costa-Gomes & Weizsacker (2008): no!
3. Ivanov (2011): no
4. Croson (2000) VCM: yes, lower contribution
5. Gachter & Renner (2010) VCM: yes, higher contribution!
6. Rutstrom & Wilcox (2009): yes. estimated parameters of a learning model vary between QSR and no elicitation
7. Healy (WP): mostly no

Survey of Experimental Results

Schotter & Trevino (2014)
Does elicitation created hedging problems across tasks?

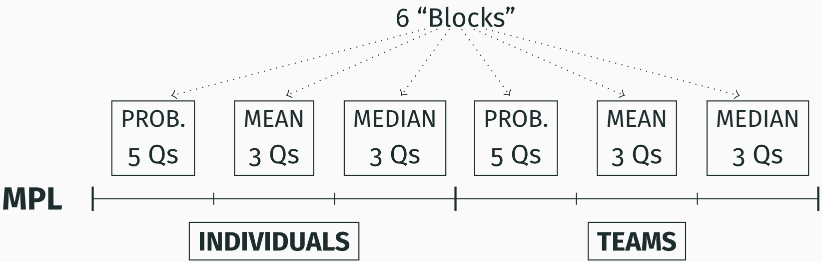
1. Blanco et al. (2010) seq PD: no
2. Armantier & Treich (2013): very little

Healy & Kagel

How to test IC of belief elicitation mechanisms?
Problem: We need to know their true belief!

- Usual technique: “Here’s a fair coin. What’s Pr(H)?”
- Problem: too suspicious!
- One solution: Bayesian updating task
- Problem: people aren’t Bayesian!
- Our idea: use team chat to look for evidence of conscious, intentional manipulation of reports
  - Subjects are in a team of two
  - Must submit the same belief report
  - Chat interface to help them coordinate
  - Do they talk about manipulating their report?
  - Do they talk about deviating from the truth?

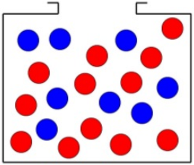
Experimental Design



- Each block has 3 or 5 questions of the same type
- Instructions before each block
- Order of blocks randomized within INDIV and TEAM
- Order of questions randomized within each block
- Three mechanisms: **MPL**, **BQSR**, **NoInfo**
  - Each subject sees only one mechanism
- INDIV first vs TEAMS first: no difference

The 11 Questions

This jar contains red and blue marbles.



The computer will randomly draw *one* marble from this jar.

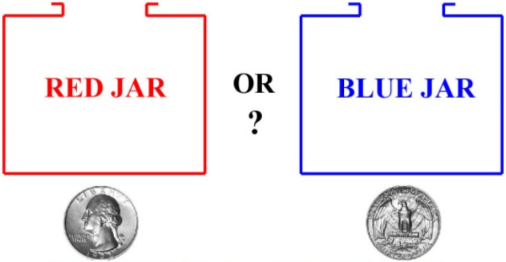
Q1: How many RED marbles are there in the jar? [ ] ($ if correct)

Q2: How many total marbles (of either color) are there in the jar? [ ] ($ if correct)

Q3: What do you think is the probability (from 0% to 100%) that a RED marble will be drawn? [ ] %

The 11 Questions

The computer will flip a coin to choose one of these two jars:

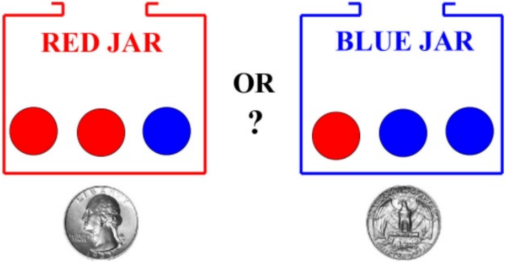


Heads: red jar is chosen. Tails: blue jar is chosen.

Q1: What do you think is the probability (from 0% to 100%) that the RED JAR was chosen? ______ %

The 11 Questions

Again, one of two jars is chosen by a coin flip. But now the jars contain 3 marbles:

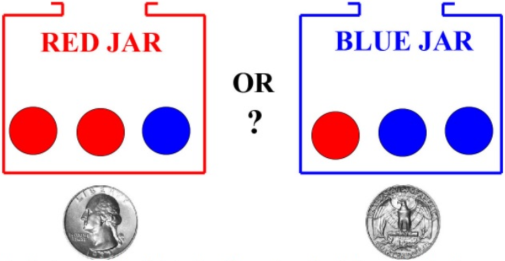


To give you a clue of which jar was chosen, we drew a marble from the chosen jar.

The marble drawn was a **BLUE** marble.

Q1: Now what do you think is the probability (from 0% to 100%) that the RED JAR was chosen? ______ %

Continuing on with the same chosen jar:



We put the first marble back into the chosen jar, shook it, and again drew a marble.

The second marble was also **BLUE**

(Thus, two **BLUE** marbles were drawn).

Q1: Now what do you think is the probability (from 0% to 100%) that the RED JAR was chosen? ______ %

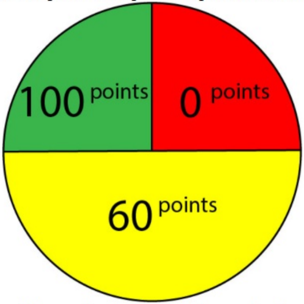
The 11 Questions

In 2005 we asked a Carnegie Mellon undergraduate this question:
What is the capital of Australia?

Q1: What do you think is the probability (from 0% to 100%) that they got this question right? ____ %

The 11 Questions

The computer will spin this spinner one time:

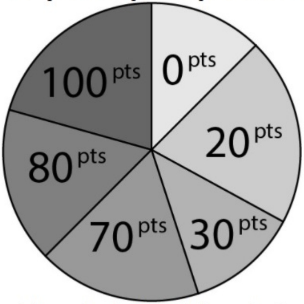


The *median* is the 'middle number.'

If the median is M, then you have ≥50% chance of getting ≥M points, *and* ≥50% chance of getting ≤M points.

Q1: I think the median # of points for this spinner is ☐ pts

The computer will spin this spinner one time:



The *median* is the 'middle number.'

If the median is M, then you have ≥50% chance of getting ≥M points, *and* ≥50% chance of getting ≤M points.

Q1: I think the median # of points for this spinner is ☐ pts

The 11 Questions

In 2005 we gave a Carnegie Mellon undergraduate student this quiz:

1. Who is credited with inventing the wristwatch in 1904?
2. Laudanum is a form of what drug?
3. The psychoactive ingredient in marijuana is THC. What does THC stand for?
4. What chemical element has the atomic number five?
5. The study of the structural and functional changes in cells, tissues and organs that underlie disease is called what?
6. What does the suffix -itis mean?
7. The bilby, bandicoot, and quokka are all representatives of what mammalian subclass?
8. Which one of the 50 United States is the only one never to have experienced an earthquake?
9. What evolutionary biologists wrote: *'Creation science' has not entered the curriculum for a reason so simple and so basic that we often mention it: because it is false.*?
10. What is the single most diverse phylum within the animal kingdom?

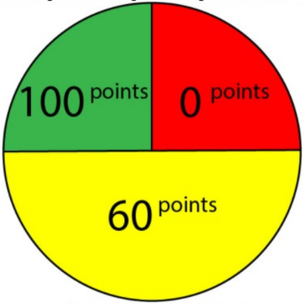
Each question was worth 10 points, for a total of 100.
The *median* is the 'middle number.'
If the median is M, then you have ≥50% chance of getting ≥M points, *and* ≥50% chance of getting ≤M poi

Q1: I think the median score for this person (from 0 to 100) is

[ ] pts

The 11 Questions

The computer will spin this spinner one time:

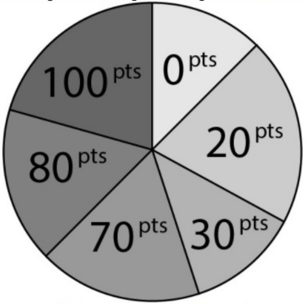


The *mean* is the 'avearge.'
If you multiply each number by its probability and add them up, you get the mean.

Q1: I think the mean # of points for this spinner is [   ] pts

The 11 Questions

The computer will spin this spinner one time:



The *mean* is the 'avearge.'
If you multiply each number by its probability and add them up, you get the mean.

Q1: I think the mean # of points for this spinner is ☐ pts

The 11 Questions

In 2005 we gave a Carnegie Mellon undergraduate student this quiz:

1. Who is credited with inventing the wristwatch in 1904?
2. Laudanum is a form of what drug?
3. The psychoactive ingredient in marijuana is THC. What does THC stand for?
4. What chemical element has the atomic number five?
5. The study of the structural and functional changes in cells, tissues and organs that underlie disease is called what?
6. What does the suffix -itis mean?
7. The bilby, bandicoot, and quokka are all representatives of what mammalian subclass?
8. Which one of the 50 United States is the only one never to have experienced an earthquake?
9. What evolutionary biologists wrote: *'Creation science' has not entered the curriculum for a reason so simple and so basic that we often mention it: because it is false.*?
10. What is the single most diverse phylum within the animal kingdom?

Each question was worth 10 points, for a total of 100.
The *mean* of their score is the 'avearge.'
If you multiply each possible score by the probability they got that score and add them up, you get the mea

Q1: I think the mean of their score (from 0 to 100) is pts

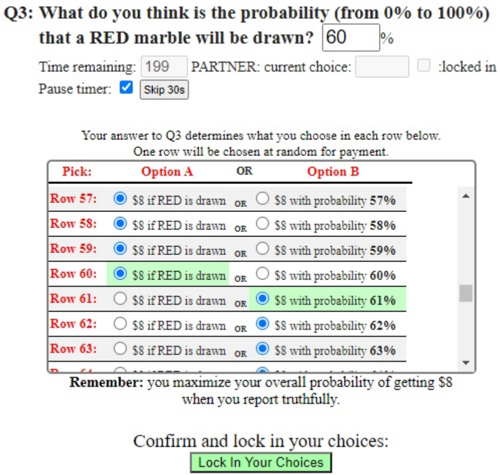
How To Present the Mechanisms

"In the first place, the subject must understand the scoring rule... This is an important reason to present the rule through some vivid tabular or graphic device..."

–Savage (1971)

- **BSR:** Wilson & Vespa (2019), Danz, Wilson & Vesterlund (2022)
- **MPL:** Holt & Smith (2016), Healy (2018)

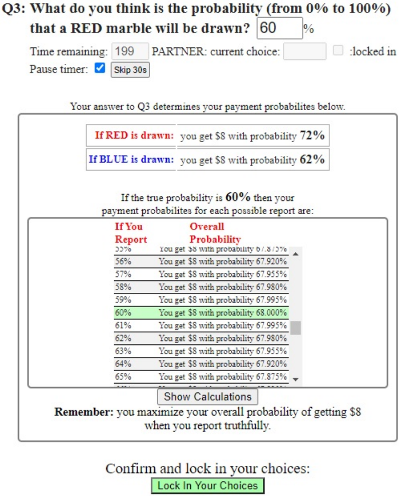
The Mechanism Interfaces: MPL



Link

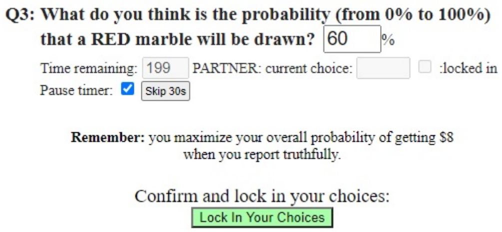
Note: subjects saw the same phrase in all three treatments

The Mechanism Interfaces: BSR



Link

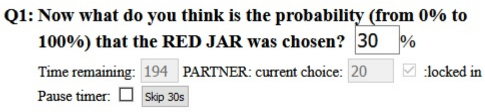
The Mechanism Interfaces: NoInfo

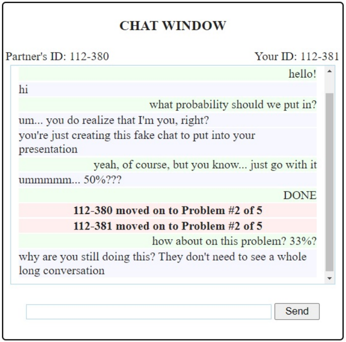


Link

Note: subjects saw the same phrase in all three treatments

Teams Interface





- Use chat window to communicate
- Must lock in the same number to proceed
- Can unlock & change ⇒ “Silent agreement”
- If time runs out, one choice is randomly used

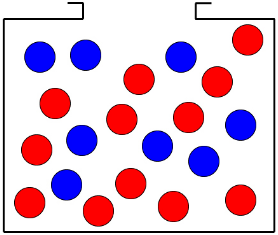
Logistics

- Usual OSU subject pool (ORSEE)
- Zoom meeting
- Less control of software environment $\Rightarrow$ missing observations
  - INDIV: 1.7% TEAM: 8.3%
- Venmo payments (option for in-person)
- \$12 show-up + possible \$8 "bonus." (59% won the bonus)

Subjects:

Mechanism:	**MPL**	**BSR**	**NoInfo**
INDIV First:	68	68	63
TEAMS First:	54	54	0
Pooled:	122	122	63

Objective-Easy #1: % Correct



Pr(Red) = 12/20 = 60%

% Correct:

	MPL	BSR	NoInfo
INDIV:	91.7%	96.6%	92.1%
TEAM:	94.8%	100%	96.4%

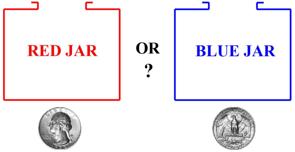
MPL seems a little worse. Are they trying to manipulate?

Objective-Easy #1: Chats

ID#181	MPL	ID#187
i have 12 for red and 8 for blue		
		12, 20, and 75%? yes
75 sounds good with me		
12|20|75%		**12|20|75%**

ID#289	MPL	ID#295
sorry I put wrong answer for 3		
12|20|50%		**12|20|50%**

Objective-Easy #2: % Correct



Pr(Red) = 50%

% Correct:

	MPL	BSR	NoInfo
INDIV:	91.5%	84.8%	93.7%
TEAM:	98.3%	93.1%	100%

Now BSR seems a little worse?

Objective-Easy #2: Chats

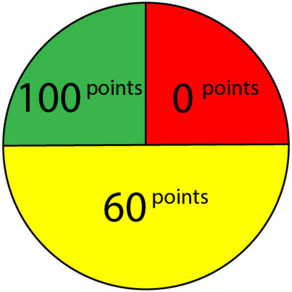
ID#390	MPL	ID#391
		50%
so theoretically it's 50 right but i think i said 48 last time just bc I'm in stats rn and we just did probability stuff about how smaller sample sizes are further from the probability so flipping it once might be 60-40 but 100 times is closer to 50-50 but ya I'm good w just 50		
		makes sense should we do 49%
sure		
49%		**49%**

Objective-Easy #2: Chats

ID#257	BSR	ID#260
		50 ?
id say 60		
		Why
cause heads is always more likely		
		Thats just false
55 is a compromise		
		Which is also wrong but whatever
55%		**55%**

ID#357	BSR	ID#365
	(no chat)	
75%		**75%**

Objective-Easy #3: % Correct



Median = 60pts

% Correct:

	MPL	BSR	NoInfo
INDIV:	69.2%	83.9%	74.2%
TEAM:	74.6%	86.1%	92.6%

Objective-Easy #3: Chats

ID#343	MPL	ID#345
well if it was 100, 0 and 50 the median would be 50 but its 60 and so id go w like 55?		
		yeah
55%		55%

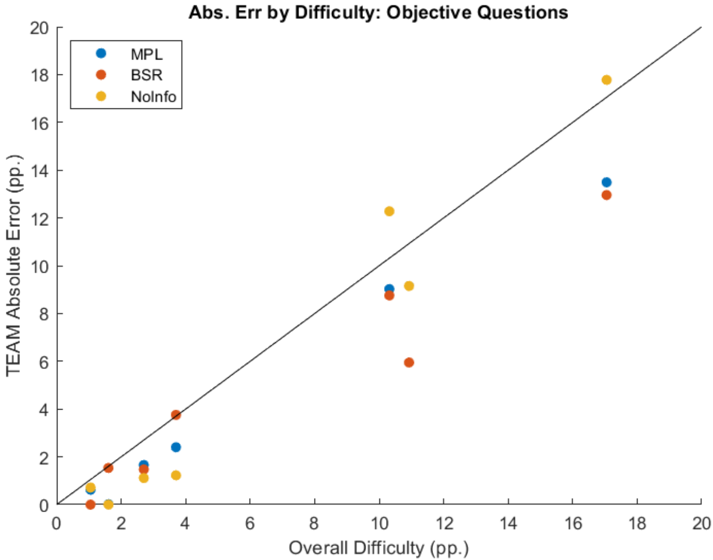
ID#352	MPL	ID#353
		I did 60
55		
		55 is good
55%		55%

Objective-Easy #3: Chats

ID#197	BSR	ID#202
		what do u think
hmm i don't remember what i said but maybe like 75? i'm not sure at all		
		love it
75%		75%

ID#302	BSR	ID#308
80?		
		yeah
80%		80%

Absolute Error by Treatment



Chat Encoding

Two Types of Evidence of IC Failures:

Calculate Playing with the calculator

- May not end up deviating from their belief

Deviate Deviate from stated belief

- May not specify *why* they're deviating

Two independent chat encoders

Chat Encoding

Two Types of Evidence of IC Failures:

Calculate Playing with the calculator

- May not end up deviating from their belief

Deviate Deviate from stated belief

- May not specify *why* they're deviating

Team-level data:

Mechanism:	MPL	BSR	NoInfo
Calculate	1	10	0
Deviate	1	1	0
Both	0	1	0

Chat Encoding

Two Types of Evidence of IC Failures:

Calculate Playing with the calculator

- May not end up deviating from their belief

Deviate Deviate from stated belief

- May not specify *why* they're deviating

Question-level data:

Mechanism:	**MPL**			**BSR**			**NoInfo**
Question:	Obj-E	Obj-H	Subj	Obj-E	Obj-H	Subj	All
Calculate	0	0	1	1	4	10	0
Deviate	1	0	0	0	0	1	0
Both	0	0	0	0	0	1	0

Subjects use the BSR calculator when clueless!

Calculate & Deviate: BSR

Capital of Australia

ID#591	BSR	ID#599
i said 90 bc Carnegie is a prestigious school and theyre smart kiddos so they hv to know this easy answer		
what do u think		
should we go higher than 90		
		I think we should go higher
95/ 100?		
95? 100? **		
		seems 100 gets the higher probability
yea with 55.9		
		**highest
		should we do 100
yes		
100		**100**

Deviate: MPL

Mean of Easy Spinner

ID#181	MPL	ID#187
the mean is 50 but i think we should do 60		
sound good with me		
i going say 60 lol		
60		60

Not Flagged: MPL

12/20/60%

ID#352	MPL	ID#353
		60%
12 red marbles, 20 total, so 60%		
		Yea but I am thinking should we really put the correct number for probability
I mean yeah i think Although its random, its the best “odds” then		
		alright
60%		**60%**

Calculate: BSR

Capital of Australia

ID#407	BSR	ID#414
hi		
		hi
		i noticed that the higher you make their percentage, the higher our probability percentage gets
yeah that's true		
		but the closer to 50, the more equal the probs
i say we go for a big one		
85		**85**

Calculate: BSR

Mean of Hard Quiz Score

ID#298	BSR	ID#312
it sounds like 50 but if i took this test i might get 3/4 right		
it looks like pretty much any number i type in i get 51/5% 50 is fine ig		
its the same no matter what we type is what ive seen		
50		**50**

$(X = M \Rightarrow 51.5\%)$

Calculate: BSR

Mean of Hard Quiz Score

ID#299	BSR	ID#303
40 technically gives the best odds		
ok		
40		**40**

Calculate: BSR

Capital of Australia

ID#359	BSR	ID#362
this was one i wasnt sure		
i originally thought a high number		
i put 90% but idk		
		i did 48 last time but we can jack up one of the probabilities id do 90
Isnt it Syndey? that is pretty well known right?		
		because it gives us 55% chance of getting red and yes it is sydney everyone knows that because of finding nemo lol
90		**90**

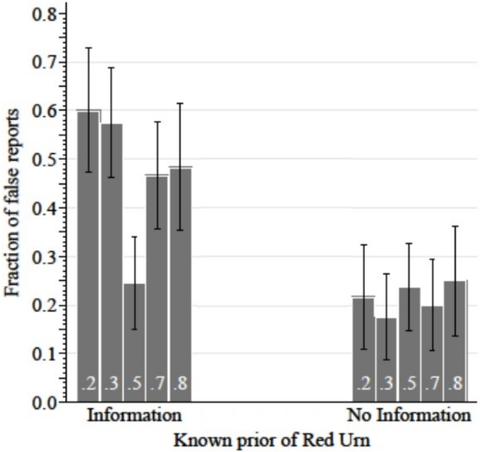
(90% ⇒ Right: 55%, Wrong: 15%)

The Story

- NoInfo performs just as well when easy, worst when hard
- Chats conclude they're **not** successfully manipulating
  - Maybe slightly more *attempts* in BSR?
- Implication: Mechanism details can be distracting **or** useful
  - Easy problems: details get in the way, $\uparrow$ mistakes
  - Harder problems: details maybe help focus, $\downarrow$ mistakes

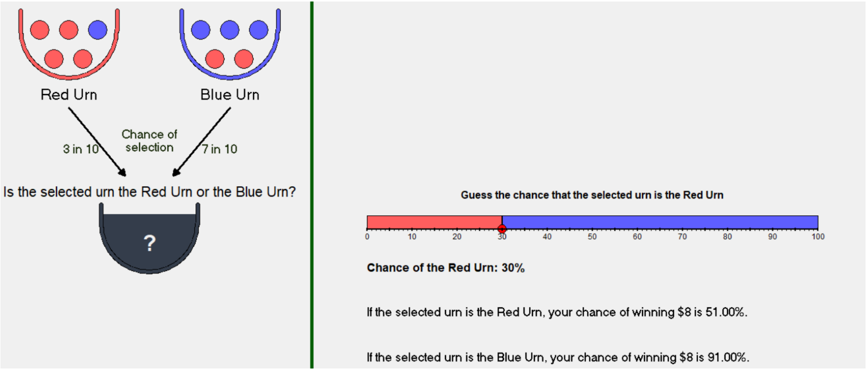
The Pittsburgh Paper

Easy Task misreport %:



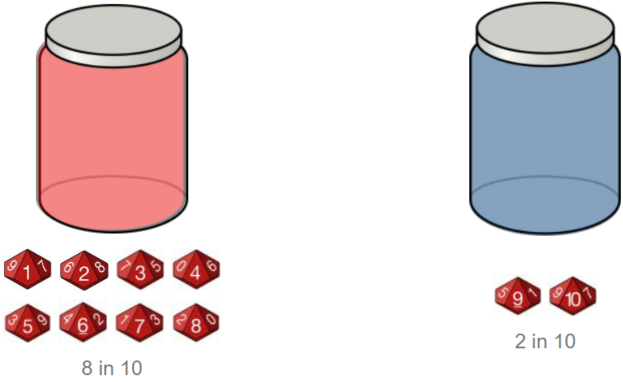
- We had $< 10\%$ at 0.5 and 0.6
- Why do they see misreporting & pull-to-center???

Danz Et Al. Choice Interface



- Clickable slider ⇒ inexact answers ⇒ pull to center??
- True probability too small??
  - Changes on every screen
  - More susceptible to distraction by payment info?

Our Choice Interface: NoInfo

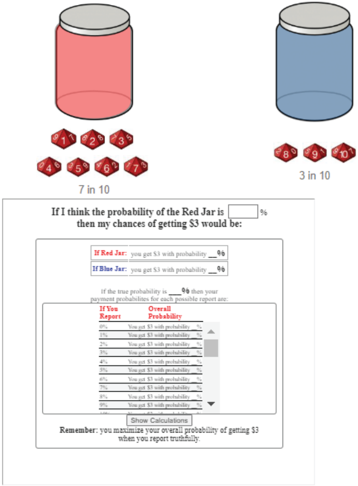


What do you think is the probability (from 0% to 100%) that the RED JAR was chosen?

[ ] %

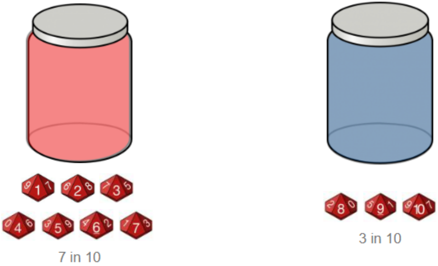
Our Choice Interface: BQSR

The computer will roll a 10-sided die to choose one of these two jars. The Red Jar is chosen if the die comes up 1 through 7.



What do you think is the probability (from 0% to 100%) that the RED JAR was chosen?

Our Choice Interface: MPL



If I think the probability of the Red Jar is ____ %
then my choices would be:

Pick:	Option A	OR	Option B
Row 0:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **0%**
Row 1:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **1%**
Row 2:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **2%**
Row 3:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **3%**
Row 4:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **4%**
Row 5:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **5%**
Row 6:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **6%**
Row 7:	◯ $3 if the Red Jar is chosen	OR	◯ $3 with probability **7%**

Remember: you maximize your overall probability of getting $3 when you report truthfully.

What do you think is the probability (from 0% to 100%) that the RED JAR was chosen?

Instructions Only

"Instructions-Only" Treatment

How I would *actually* do elicitation:

- Mechanism details in Sinstructions
- No details on decision screens

Details

Prolific + Qualtrics

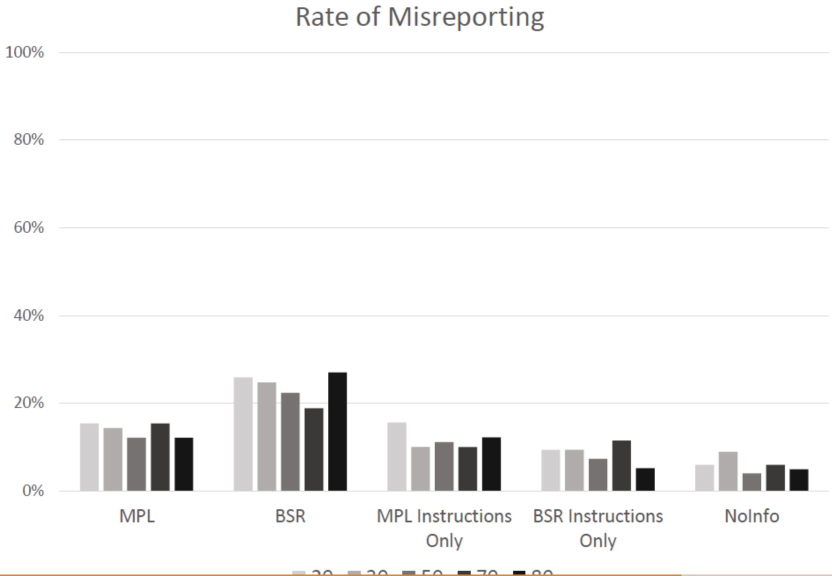
US adults 18+

3 comprehension Q's

	Total n	% Pass Comp. Test
MPL	99	92%
BQSR	99	86%
MPL-InstrOnly	100	90%
BQSR-InstrOnly	101	95%
NoInfo	103	98%

χ^2 test p-value: 0.015

Robust Replication Results



Differences?

"Robust replication" vs. "exact replication"

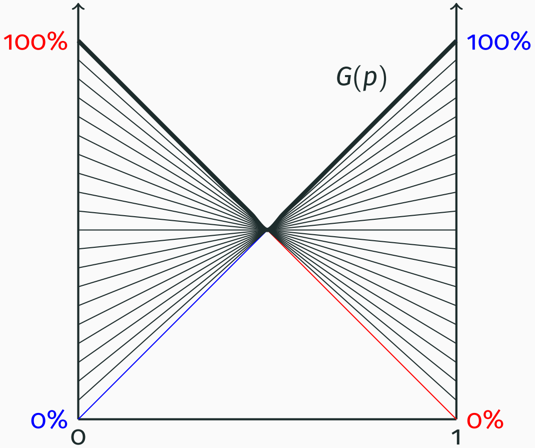
Differences:

1. Pitt Lab adults vs. Prolific US adults
2. Clickable slider vs. text input
3. Different illustrations of the question
4. We scaled BQSR to make expected payment = MPL
5. Instructions similar, not the same
6. Different calculator interfaces

⋮

A Non-IC Mechanism

Recall Linear Scoring Rule (LSR):



$S(q, 0) = 1 - q$ $\quad\quad\quad\quad$ $S(q, 1) = q$

$q^* = 0$ if $p < 50$, $q^* = 100$ if $p > 50$

A Non-IC Mechanism

Why test this?

1. Validating the chat methodology
    - They *should* deviate...
    - so do we see them chat about it?
2. Does incentive compatibility even matter?
    - Maybe they don't pay attention!

A Non-IC Mechanism

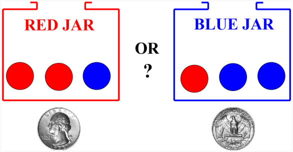
Preliminary results:

- Chat data:
  - Out of 30+ subjects, only **one** mentions it
  - And their partner dismisses it!
- Choice data:
  - INDIV: a few more cases of 100 and 0!!
  - TEAM: no differences

I can't get people to lie!!!
Really don't replicate Danz et al. (2022)

Tangential Results

Errors in Bayesian Updating



- One Blue Draw:
  - $Pr(R|b) = Pr(R) * Pr(b|R)$. 17%
  - Marble draw is uninformative. 50%
- Two Blue Draws:
  - $Pr(R|bb) = Pr(R) * Pr(b|R) * Pr(b|R)$. 6%
  - Second draw gives no new info. Same as one.
  - Marble draws are uninformative. 50%
  - Second draw was with replacement. 0%

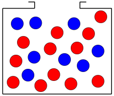
Does The Truth Win?

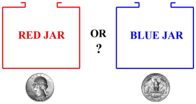
“Truth-Wins” Norm:

2 Right: Both players were correct in INDIV

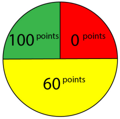
1 Right: One player was correct in INDIV

Team Right: Both players correct in TEAM ($n = 73$ teams)



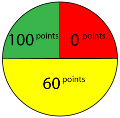


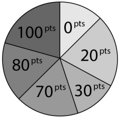
Median

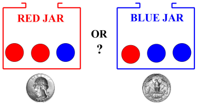


			Median
Team Right|2 Right:	80/83	64/69	46/51
Team Right|1 Right:	8/10	22/24	26/34
Team Right|0 Right:	0/1	1/1	1/9

Does The Truth Win?







	Mean	Median	1 BLUE
Team Right|2 Right:	26/29	16/21	7/8
Team Right|1 Right:	29/42	24/41	26/47
Team Right|0 Right:	6/23	8/32	3/39

Aggregating Beliefs

1. Prediction Markets
   - Double-auction w/ Arrow securities
2. Market Scoring Rules
3. Parimutuel Betting Markets
4. The Delphi Method
5. Bayesian Truth Serum

Prediction Markets

- Double auction w/ Arrow securities (\$1 if E)
- Wave of popularity: Wolfers and Zitzewitz [2004]
  - Iowa Electronic Markets [Berg et al., 1996]
  - TradeSports & InTrade
  - In-house markets
    - Google [Cowgill et al., 2009]
    - HP [Ho and Chen, 2007]
  - DARPA Policy Analysis Market [Hanson, 2007]
- Theory problem: does Walrasian equilibrium really aggregate info?
  - Manski [2006]: No
  - Other models: yes (cite needed!)

Market Scoring Rules

Hanson [2003], Ledyard et al. [2009]

- Start with public distribution p_0
- Player i moves it to some p_1
- Paid $S(p_1, x) - S(p_0, x)$
- IC since $S(p_0, x)$ doesn't depend on p_1
    - Except for dynamic incentives...
- Player i "buys out" previous player

Pari-Mutuel Betting

- Bettor i bets b_{ij} on horse j
- If horse k wins, bettor i gets

$$\underbrace{\left(\sum_{ij} b_{ij} - T\right)}_{\text{net proceeds after take } T} \underbrace{\frac{b_{ik}}{\sum_{\iota} b_{\iota k}}}_{i\text{'s bet share on } k}$$

- Koessler et al. [2002]: fully-revealing BNE if simultaneous, not seq.
- Behavioral observations:
  - Mirages: Camerer and Weigelt [1991]
  - Favorite-Longshot Bias: Snowberg and Wolfers [2006]
  - End-Of-Day Risk Seeking (Camerer?)

Iterated Polls/Delphi Method

Simple procedure:

1. Privately ask everyone's prior
2. Reveal all priors (or aggregate) to everyone
3. Players update
4. Repeat *m* times (or until convergence)
5. Pay everyone via scoring rule for final *p*

- Naive play gives info aggregation
- Dynamic incentives? McKelvey and Page [1990]
    - "Last moves" are incentive compatible

An Experimental Test

Healy et al. [2010]

- Compare DA, MktSR, Parimutuel, & Poll
- Thin markets: $n = 3$.
- $|\Omega| = 2$ vs. $|\Omega| = 8$, Traders see different # of signals

Signal structure (common info):

Table 1 Distribution f for the Two-State Experiments

θ	$f(\theta)$	$f(H \mid \theta)$	$f(T \mid \theta)$
X	1/3	0.2	0.8
Y	2/3	0.4	0.6

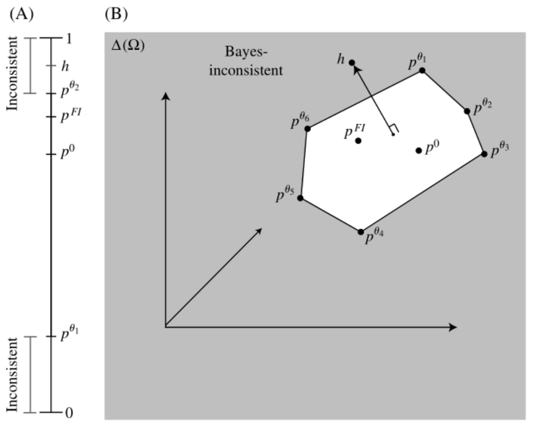
Table 2 Distribution f for the Eight-State Experiments

θ	$f(\theta)$	TTT	TTH	THT	THH	HTT	HTH	HHT	HHH
XYZ	1/6	0.320	0.213	0.160	0.107	0.040	0.027	0.080	0.053
XZY	1/6	0.320	0.160	0.213	0.107	0.040	0.080	0.027	0.053
YXZ	1/6	0.320	0.213	0.040	0.027	0.160	0.107	0.080	0.053
YZX	1/6	0.320	0.040	0.213	0.027	0.160	0.080	0.107	0.053
ZXY	1/6	0.320	0.160	0.040	0.080	0.213	0.107	0.027	0.053
ZYX	1/6	0.320	0.040	0.160	0.080	0.213	0.027	0.107	0.053

An Experimental Test

Measures of Performance:

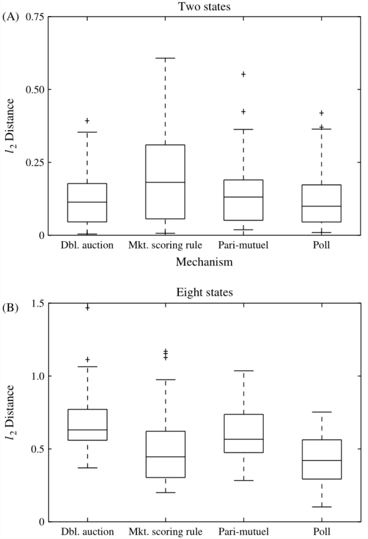
Figure 2 Bayes-Inconsistent Outcomes with (A) Two States and (B) More Than Two States



1. l_2 distance from "full info posterior"
2. Bayes-Inconsistency

An Experimental Test

Distance to full-info posterior:



An Experimental Test

Distance to Bayes-consistency ($|\Omega| = 8$):

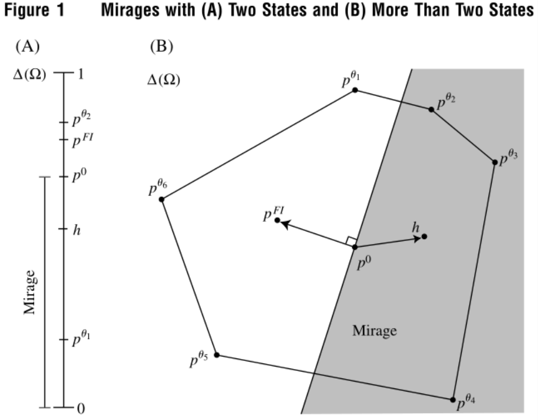
Table 10 ***p*-Values of Mechanism-by-Mechanism Wilcoxon Tests Comparing the Severity of Bayes-Inconsistency, as Measured by the Distance Between the Mechanism Output Distribution and the Convex Hull of the Limit Posteriors**

Eight states	Avg. dist	Dbl. auction	Mkt. scoring rule	Pari-mutuel	Poll
Avg. distance	—	0.447	0.362	0.398	0.312
Dbl. auction	0.447	—	**0.001**	0.107	$<$ **0.001**
Mkt. scoring rule	0.362	—	—	0.180	0.257
Pari-mutuel	0.398	—	—	—	**0.008**
Poll	0.312	—	—	—	—

Note. 10% Significance ordering: DblAuc $\succeq$ Pari $\succeq$ MSR $\succeq$ Poll, DblAuc $\succ$ MSR $\succeq$ Poll, DblAuc $\succeq$ Pari $\succ$ Poll.

An Experimental Test

Measures of Performance:



3. Mirages
4. No trade!

An Experimental Test

Mirages and No Trade ($|\Omega| = 8$):

Table 7 **Number of Periods in Each Session (Out of 8) and Number of Periods Total (Out of 32) in Which Each Type of Catastrophic Failure Occurs in the Eight-State Experiments**

	Dbl. auction		Mkt. scoring rule		Pari-mutuel		Poll	
	(S5, S6, S7, S8)	Tot.	(S3, S4, S1, S2)	Tot.	(S1, S2, S3, S4)	Tot.	(S7, S8, S5, S6)	Tot.
No trade	(0, 0, 0, 0)	0	(0, 0, 0, 0)	0	(0, 0, 8, 1)	9	(0, 0, 0, 0)	0
Mirage	(3, 1, 4, 4)	12	(1, 1, 2, 3)	7	(3, 1, 0, 3)	7	(0, 1, 2, 0)	3
None	(5, 7, 4, 4)	20	(7, 7, 6, 5)	25	(5, 7, 0, 4)	16	(8, 7, 6, 8)	29

Note. Every mechanism is Bayes-inconsistent in every period.

An Experimental Test

Summary:

Table 11 Summary of Results

	Two states				Eight states			
Summary	Error	No trade	Mirage	Inconsistent	Error	No trade	Mirage	Inconsistent
Dbl. auction	✓	✓	✓	✓	×	✓	×	×
MSR	×*	✓	✓	✓	✓	✓	✓	✓
Pari-mutuel	✓	×*	✓	✓	×	×*	✓	×
Poll	✓	✓	✓	×*	✓	✓	✓	✓

Notes. A ✓ indicates the mechanism was not significantly outperformed by some other mechanism in that measure and an × indicates that it was. An ×* denotes either marginal significance (all *p*-values less than but close to 0.10) or cases where proper statistical tests were unavailable.

Bayesian Truth Serum

Prelec [2004]
Method to get truthful answers to a survey question.

- Agents: $i \in \{1, \dots, n\}$.
- Options/answers: $j \in \{1, \dots, m\}$
- Each i announces:
    1. their answer $t_i \in \{1, \dots, m\}$
    2. their distribution of other's answers $p_i(\cdot) \in \Delta(\{1, \dots, m\})$
- Define:
    - $I_{ij} = 1$ iff $t_i = j$
    - $\bar{x}_j = \frac{1}{n} \sum_i I_{ij}$
      Actual average frequency of j
    - $\bar{y}_j = \exp\left(\frac{1}{n} \sum_i \log(p_i(j))\right)$
      Geometric average predicted frequency of j

Bayesian Truth Serum

Incentives:

- “info score” for each option: $\iota(j) = \log\left(\frac{\bar{x}_j}{\bar{y}_j}\right)$
- prediction penalty: $\rho(p_i) = \sum_{j=1}^{m} \bar{x}_j \log\left(\frac{p_i(j)}{\bar{x}_j}\right)$

Payoff:

$$\pi(t_i, p_i(\cdot)) = \iota(j) + \alpha\rho(p_i)$$

Theorem: Assume opinions (t_i) are exchangeable and n is large. Then truth-telling is a Bayes-Nash equilibrium. Furthermore, among equilibria, it is the equilibrium that maximizes the expected info score

Literatur

Joyce Berg, Robert Forsythe, and Thomas Rietz. What makes markets predict well? Evidence from the Iowa Electronic Markets. In *Understanding Strategic Interaction: Essays in Honor of Reinhard Selten*. Springer, New York, 1996.

Colin F. Camerer and Keith Weigelt. Information Mirages in Experimental Asset Markets. *Journal of Business*, 64(4):463–493, 1991.

Bo Cowgill, Justin Wolfers, and Eric Zitzewitz. Using Prediction Markets to Track Information Flows: Evidence from Google. 2009.

Robin Hanson. Combinatorial Information Market Design. *Information Systems Frontiers*, 5(1):107–119, 2003.

Robin Hanson. The Policy Analysis Market: A Thwarted Experiment in the Use of Prediction Markest for Public Policy. *Innovations*, 2(3):73–88, 2007.

Paul J. Healy, Sera Linardi, J. Richard Lowery, and John O. Ledyard. Prediction Markets: Alternative Mechanisms for Complex Environments with Few Traders. *Management Science*, 56(11):1977–1996, November 2010. ISSN 0025-1909. doi: 10.1287/mnsc.1100.1226.

Teck-Hua Ho and Kay-Yut Chen. New Product Blockbusters: The Magic and Science of Prediction Markets. *California Management Review*, 50(1): 144–158, October 2007. ISSN 0008-1256, 2162-8564. doi: 10.2307/41166420.

Frederic Koessler, Charles Noussair, and Anthony Ziegelmeyer. Parimutuel Betting under Asymmetric Information. *Journal of Mathematical Economics*, 44:733–744, 2002.

John O. Ledyard, Robin D. Hanson, and Takashi Ishikida. An Experimental Test of Combinatorial Information Markets. *Journal of Economic Behavior and Organization*, 69:182–189, 2009.

References: iii

Charles F. Manski. Interpreting The Predictions of Predictions Markets. *Economics Letters*, 91(3):425–429, 2006.

Richard D. McKelvey and Talbot Page. Public and Private Information: An Experimental Study of Information Pooling. *Econometrica*, 58(6): 1321–1339, 1990.

Dražen Prelec. A Bayesian truth serum for subjective data. *Science (New York, N.Y.)*, 306(5695):462–466, 2004.

Erik Snowberg and Justin Wolfers. Explaining the Favorite-Longshot Bias: Is it Risk-Love, or Misperceptions? October 2006.

Justin Wolfers and Eric Zitzewitz. Prediction Markets. *Journal of Economic Perspectives*, 18(2):107–126, 2004.