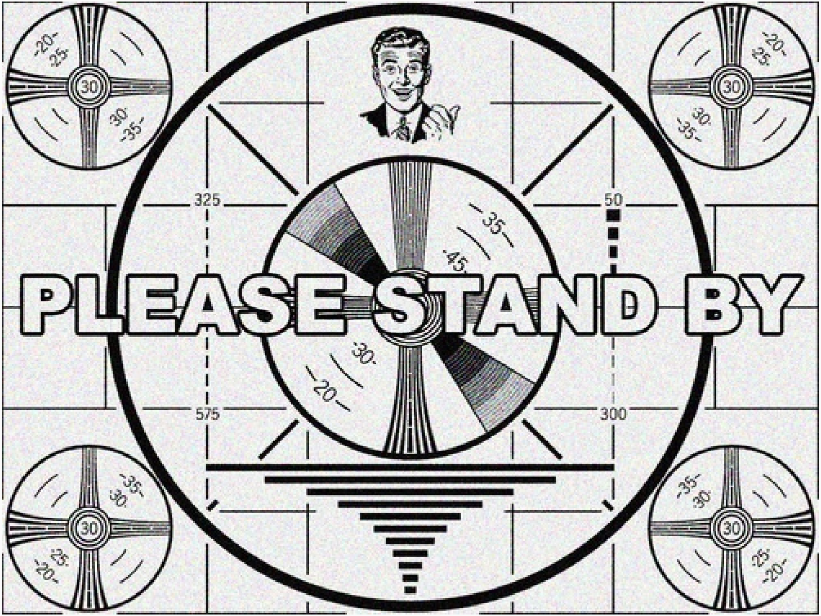

PLEASE STAND BY

Testing Elicitation Mechanisms via Team Chat

Paul J. Healy (OSU)
John Kagel (OSU)

MiddExLab

Introduction

Beliefs are central to most economic theories & predictions

Introduction

Beliefs are central to most economic theories & predictions

Therefore, it's important that we're able to measure them accurately

So Many Mechanisms!!

But how should we elicit them?

- Unincentivized
  - No: Ramsey (1931), de Finiti (1937), Savage (1954)
- Quadratic scoring rule (QSR; Brier 1950)
  - Others: Logarithmic, spherical...
  - QSR corrected for risk aversion (Harrison et al. 2014)
- Binarized scoring rules (BSR; Savage 1971, Hossain & Okui 2013)
  - "Paired-uniform" BSR (Wilson & Vespa 2017)
- BDM for probabilities (Marschak 1963, Grether 1981)
  - Clock BDM (Karni 2009)
- Multiple Price List (MPL; Holt & Smith 2016)

So Many Mechanisms!!

But how should we elicit them?

- Unincentivized
    - No: Ramsey (1931), de Finiti (1937), Savage (1954)
- Quadratic scoring rule (QSR; Brier 1950)
    - Others: Logarithmic, spherical...
    - QSR corrected for risk aversion (Harrison et al. 2014)
- Binarized scoring rules (BSR; Savage 1971, Hossain & Okui 2013)
    - "Paired-uniform" BSR (Wilson & Vespa 2017)
- BDM for probabilities (Marschak 1963, Grether 1981)
    - Clock BDM (Karni 2009)
- Multiple Price List (MPL; Holt & Smith 2016)

Each mech is IC under different assumptions.
Our focus: BSR & MPL

What Do The Data Say?

- Offerman & Sonnemans (2004): QSR$\sim$None
- Armentier & Triech (2013): QSR$\succ$None
- Huck & Weizsacker (2002): QSR$\succ$BDM
- Hollars et al. (2010): BDM$\succ$QSR
- Hao & Houser (2012): BDM-Clock$\succ$BDM
- Hossain & Okui (2013): BSR$\succ$QSR
- Harrison et al. (2014): BSR$\sim$QSR-Corr$\succ$QSR
- Wilson & Vespa (2017): BSR$\succ$PU-BSR
- Holt & Smith (2016); MPL$\succ$BDM

Our focus: BSR and MPL

Our Motivations

- Offerman & Sonnemans (2004): QSR$\sim$None
- Armentier & Triech (2013): QSR$\succ$None
- Huck & Weizsacker (2002): QSR$\succ$BDM
- Hollars et al. (2010): BDM$\succ$QSR
- Hao & Houser (2012): BDM-Clock$\succ$BDM
- Hossain & Okui (2013): BSR$\succ$QSR
- Harrison et al. (2014): BSR$\sim$QSR-Corr$\succ$QSR
- Wilson & Vespa (2017): BSR$\succ$PU-BSR
- Holt & Smith (2016); MPL$\succ$BDM

Motivation #1: Compare MPL to BSR in theory and in the lab

Our Motivations

Our theory results:

1. MPL is IC under weaker assumptions than BSRs
2. $\exists$ isomorphism between MPLs and some BSRs, but not all

Our lab results (so far):

1. Between MPL and BSR, it's basically a tie

Motivation #1: Compare MPL to BSR in theory and in the lab

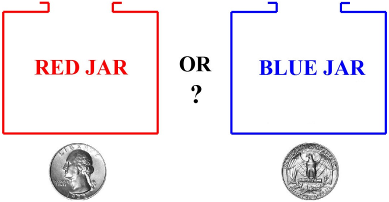
How Can You Test if an Elicitation Mechanism Works??

Motivation #2: Experiments testing elicitation are... tricky

- Need to know their belief to test whether they report truthfully
- Two methods:
    1. Coherence of subjective beliefs ($\sum_i p_i = 1$, e.g.)
    2. Induce-then-elicit objective beliefs

Example: Objective-Easy Questions

Holt & Smith (2016), Danz et al. (2020), etc.

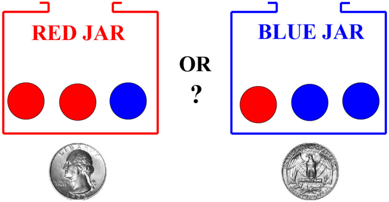


Pro: Almost certainly know their belief

Con: Too suspicious! “Deviation” might be distrust, confusion

Example: Objective-Hard Questions

Holt & Smith (2016), Danz et al. (2020), etc.



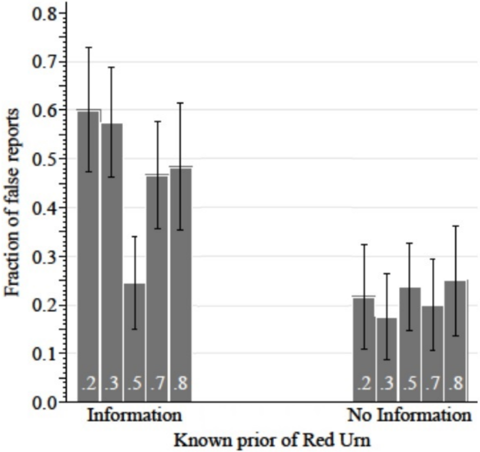
Signal: Two BLUE marbles were drawn w/ replacement

Pro: Less suspicious

Con: Too hard! "Deviation" might be confusion, errors

Danz, Vesterlund & Wilson (2020)

Objective-Easy misreport %'s:



- information $\Rightarrow$ manipulation!
- Are they *really* trying to manipulate, or are they just confused?

Our Project

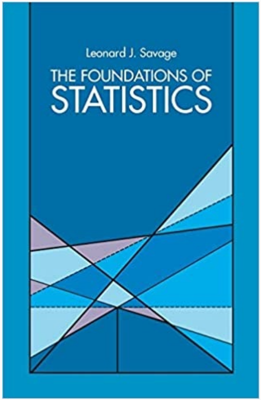
- Have subjects in teams of two, working together via chat
  - Cooper & Kagel (2005,2009,2020)
- Scan chat transcripts for (1) true beliefs, (2) manipulation
- Question: Objective-Easy, Objective-Hard, Subjective
- Compare **BSR**, **MPL**, and **NoInfo**
- Also look at eliciting means & medians

Experimental Results:

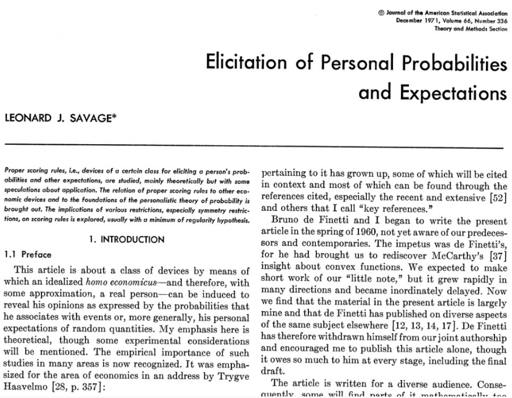
1. **NoInfo** performs best on Objective-Easy questions
   ...but worst on Objective-Hard questions
2. Very little evidence of manipulation in the chat
3. Evidence of confusion and mistakes
   ...especially when mech. details are given

Theory

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]:

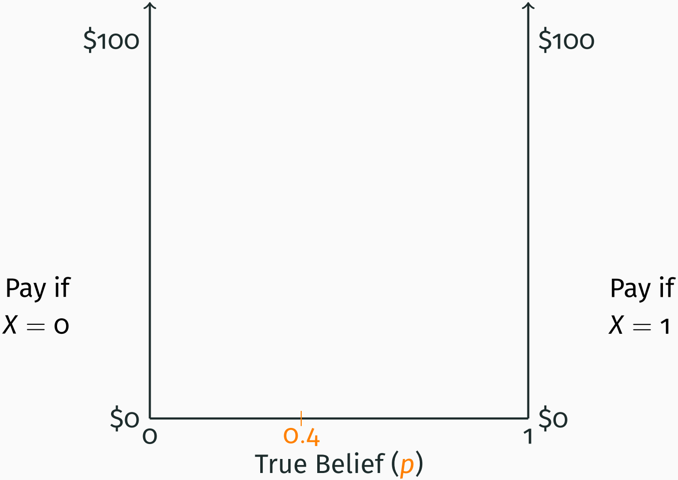
pertaining to it has grown up, some of which will be cited in context and most of which can be found through the references cited, especially the recent and extensive [52] and others that I call "key references."

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)

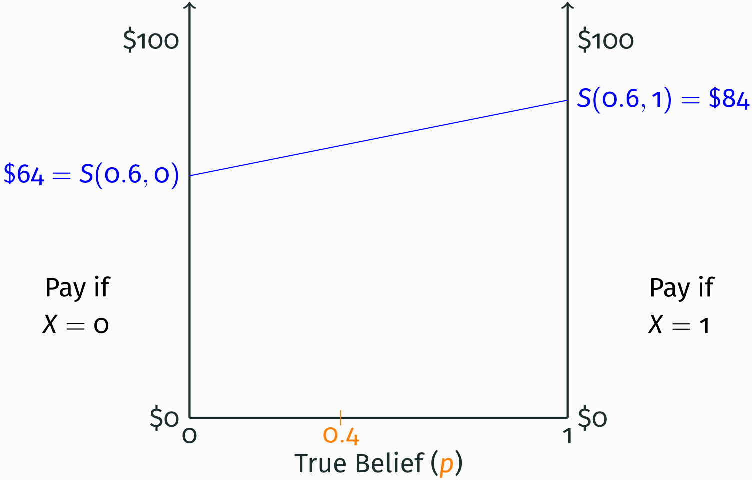
Scoring Rules (Savage 1971)



Two states: $X \in \{0, 1\}$. Announce $q = Pr(X = 1)$.

If $X = 0$, pay $S(q, 0)$. If $X = 1$, pay $S(q, 1)$.

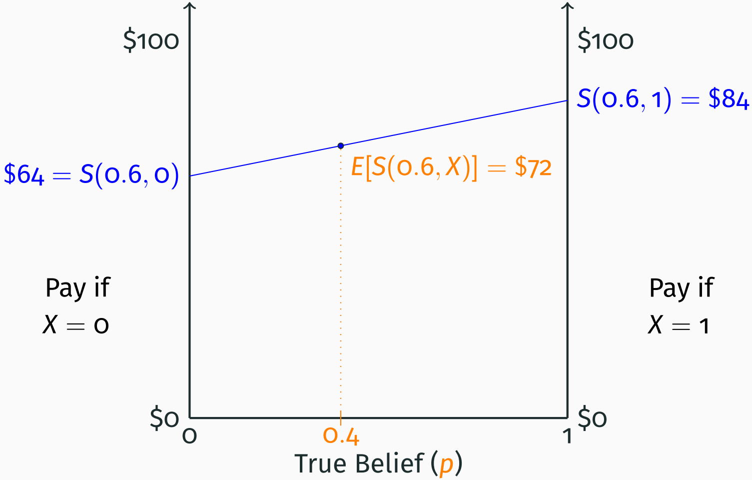
Scoring Rules (Savage 1971)



Two states: $X \in \{0, 1\}$. Announce $q = Pr(X = 1)$.

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

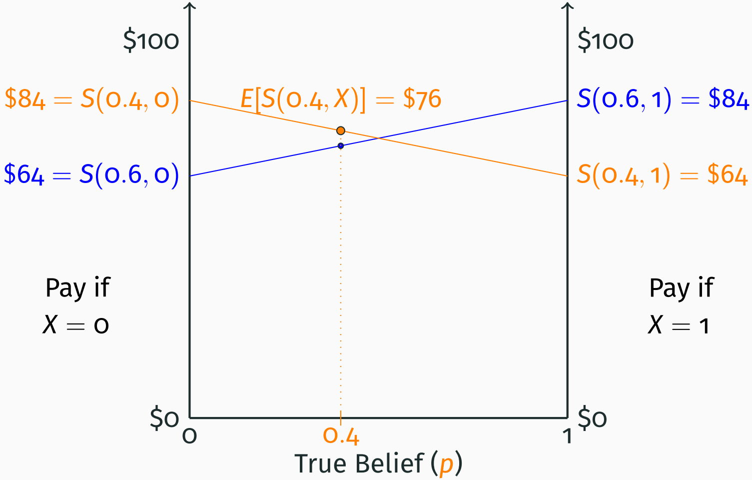
Scoring Rules (Savage 1971)



Two states: $X \in \{0, 1\}$. Announce $q = Pr(X = 1)$.

For now, assume **risk neutrality**

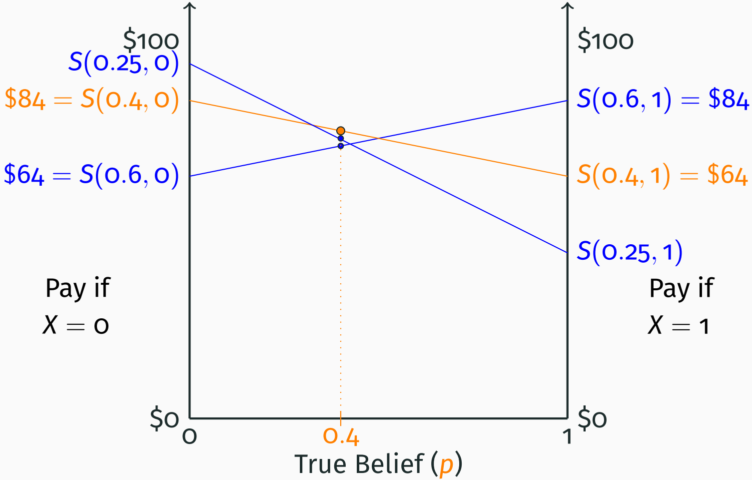
Scoring Rules (Savage 1971)



Truthful announcement ↑ E[payment]

For now, assume **risk neutrality**

Scoring Rules (Savage 1971)



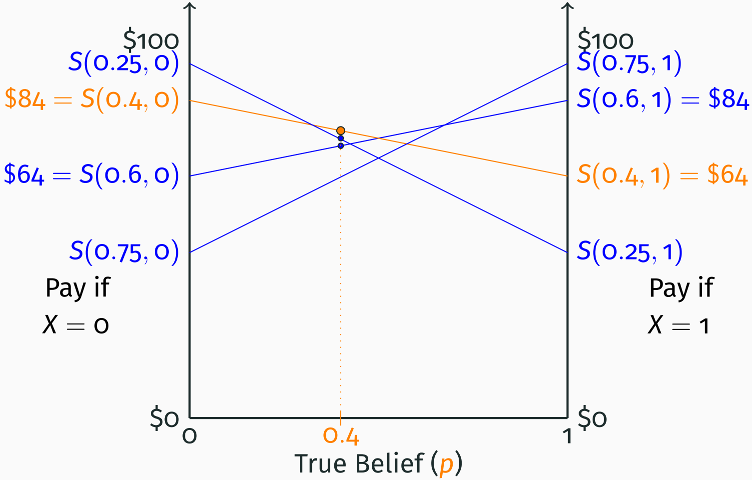
Pay if $X = 0$

Pay if $X = 1$

Any deviation $\downarrow$ E[payment]

For now, assume **risk neutrality**

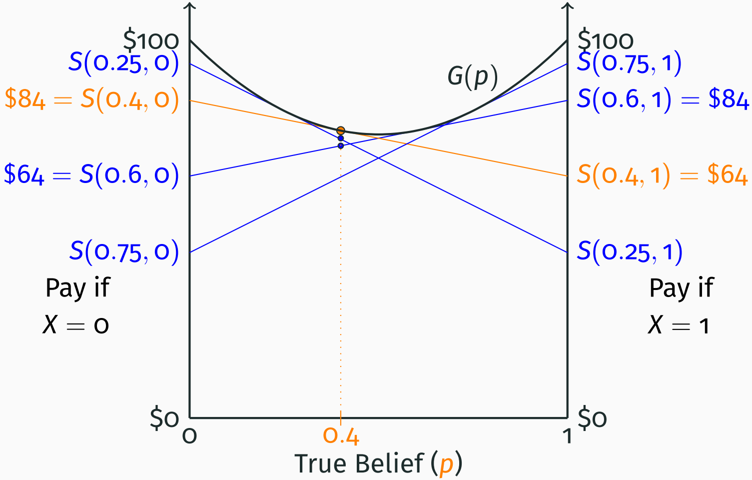
Scoring Rules (Savage 1971)



Any deviation $\downarrow$ E[payment]

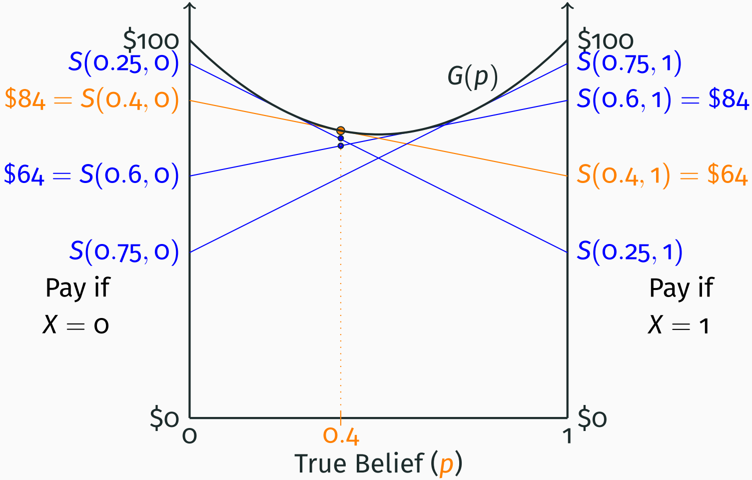
For now, assume **risk neutrality**

Scoring Rules (Savage 1971)



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

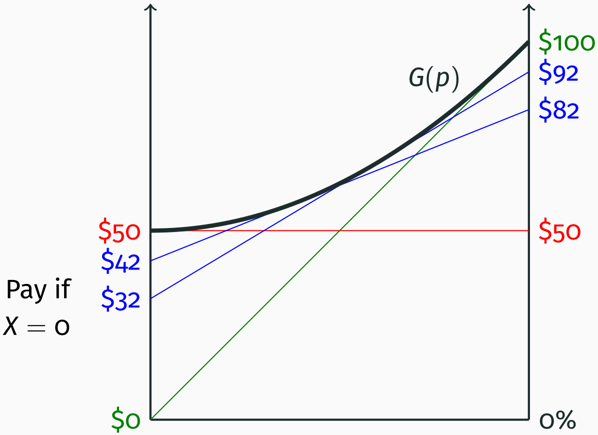
Scoring Rules (Savage 1971)



Any convex $G(p)$ will work.

Quadratic scoring rule, logarithmic, spherical...

A “Flat-To-Steep” Scoring Rule



A “flat-to-steep” scoring rule

Risk Neutrality

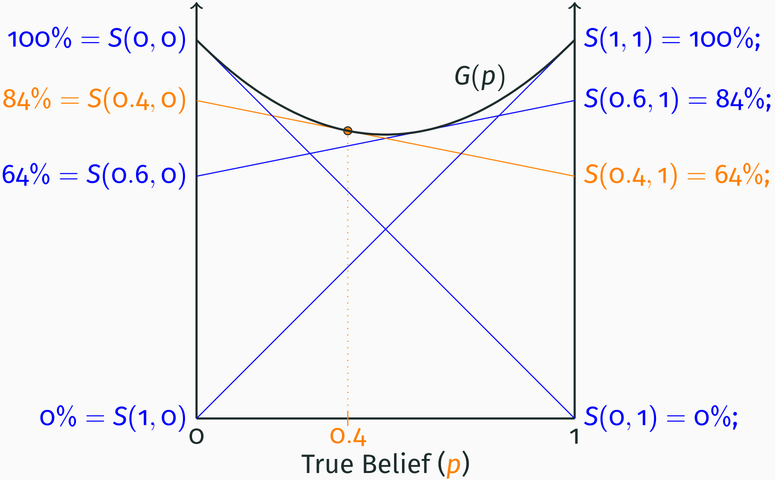
IC requires risk neutrality. Savage (1971) gives 2 solutions:

1. Pay small amounts
2. Pay in probabilities
    - Pay some % chance of winning $8
    - EU: $p \cdot u(\$8)$ is linear in p
    - Savage (1971) $\rightarrow$ C. Smith (1961) $\rightarrow$ Savage (1954)
    - "Binarized" Scoring Rules (BSR; Hossain & Okui 2013)

Does paying in probabilities work?

- In general: no (Selten et. al 1999, e.g.)
- For scoring rules: yes (Hossain & Okui 2013, e.g.)

Binarized Scoring Rules



Conditions for Incentive Compatibility

Proof of Incentive Compatibility:

$$\underbrace{p \cdot S(p,1) + (1-p) \cdot S(p,0)}_{Pr(\$8) \text{ if truth}} > \underbrace{p \cdot S(q,1) + (1-p) \cdot S(q,0)}_{Pr(\$8) \text{ if lie}}$$

This requires "Subjective-Objective Reduction"

- Weakening of ROCL
    - Applies only to binary lotteries
- Rules out perceived correlation, probability weighting, etc.

Multiple Price Lists (MPL)

Row#	Option A	OR	Option B
1	\$8 if $X = 1$	or	\$8 w/ prob 1%
2	\$8 if $X = 1$	or	\$8 w/ prob 2%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
q	\$8 if $X = 1$	or	\$8 w/ prob q%
$q + 1$	\$8 if $X = 1$	or	\$8 w/ prob $q + 1$%
$q + 2$	\$8 if $X = 1$	or	\$8 w/ prob $q + 2$%
$q + 3$	\$8 if $X = 1$	or	\$8 w/ prob $q + 3$%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	\$8 if $X = 1$	or	\$8 w/ prob 99%
100	\$8 if $X = 1$	or	\$8 w/ prob 100%

Choose Option A or Option B (single switch point q)

One row randomly selected for payment

Multiple Price Lists (MPL)

Row#	Option A	OR	Option B
1	\$8 if $X = 1$	or	\$8 w/ prob 1%
2	\$8 if $X = 1$	or	\$8 w/ prob 2%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
q	\$8 if $X = 1$	or	\$8 w/ prob q%
$q+1$	\$8 if $X = 1$	or	\$8 w/ prob $q+1$%
$q+2$	\$8 if $X = 1$	or	\$8 w/ prob $q+2$%
$q+3$	\$8 if $X = 1$	or	\$8 w/ prob $q+3$%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	\$8 if $X = 1$	or	\$8 w/ prob 99%
100	\$8 if $X = 1$	or	\$8 w/ prob 100%

"Multiple Price List" (MPL) version of BDM for probabilities

Holt & Smith (2016), Healy (2018)

Multiple Price Lists (MPL)

Row#	Option A	OR	Option B
1	\$8 if $X = 1$	or	\$8 w/ prob 1%
2	\$8 if $X = 1$	or	\$8 w/ prob 2%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
q	\$8 if $X = 1$	or	\$8 w/ prob q%
$q+1$	\$8 if $X = 1$	or	\$ w/ prob $q+1$%
$q+2$	\$8 if $X = 1$	or	\$ w/ prob $q+2$%
$q+3$	\$8 if $X = 1$	or	\$8 w/ prob $q+3$%
$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	\$8 if $X = 1$	or	\$8 w/ prob 99%
100	\$8 if $X = 1$	or	\$8 w/ prob 100%

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

I.C. as long as subject respects **statewise dominance** in rows

MPL vs BSR

BSR is I.C.

$$\Updownarrow$$

Subjective-Objective Reduction

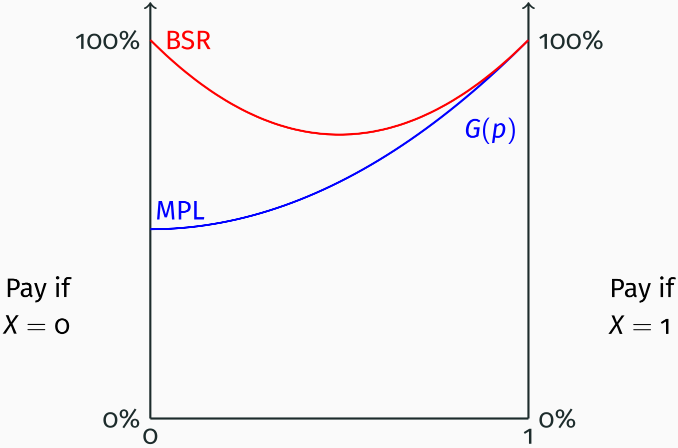
$$\Downarrow$$

Statewise Dominance

$$\Downarrow$$

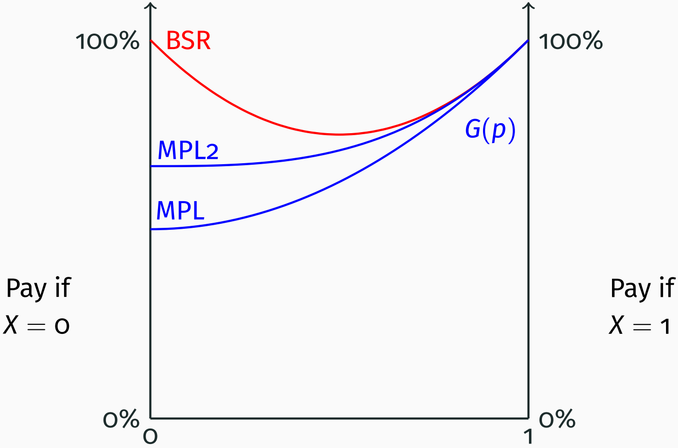
MPL is I.C.

Converting Between MPLs and BSRs



If you reduce objective lotteries in an MPL, you get a scoring rule

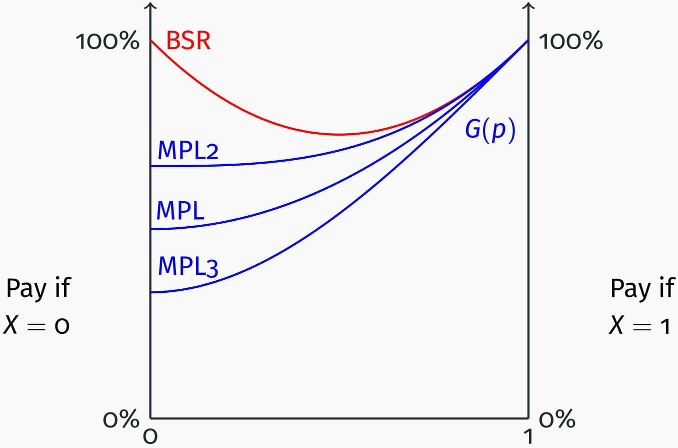
Converting Between MPLs and BSRs



If you reduce objective lotteries in an MPL, you get a scoring rule

Different row probabilities $\Rightarrow$ different $G(p)$

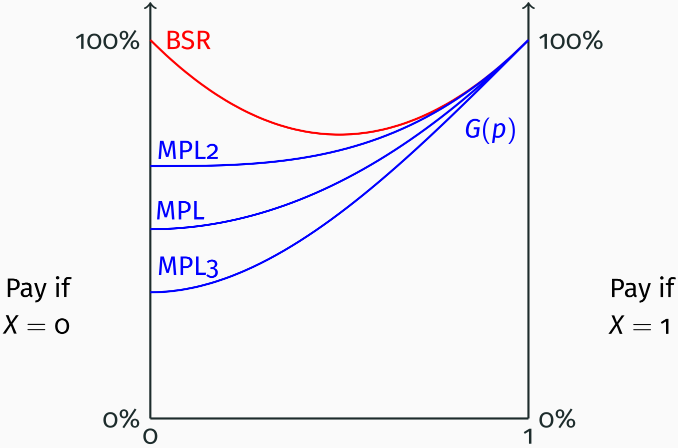
Converting Between MPLs and BSRs



If you reduce objective lotteries in an MPL, you get a scoring rule

Different row probabilities $\Rightarrow$ different $G(p)$

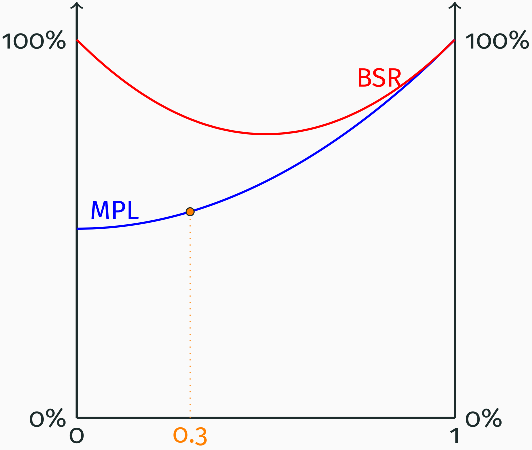
Converting Between MPLs and BSRs



Proposition: $G(p)$ is equiv. to an MPL if and only if

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

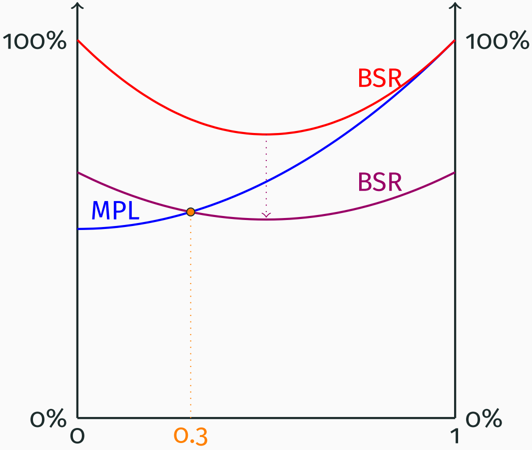
Equalizing Incentives



How to equalize incentives across scoring rules?

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

Equalizing Incentives



How to equalize incentives across scoring rules?
Shift depends on researcher's best guess of p

More Than Two States

- What if X can take more values?
    - Ex: score on a quiz, GDP next quarter
- Could elicit $Pr(X = x)$ for every possible x... but that's a lot!
- The BQSR elicits the subject's **mean** for X
    - BQSR: $S(m, x) = \left(1 - (x - m)^2\right)$
    - Still paying in probabilities (rescale X to $[0, 1]$)
    - Still requiring S-O Reduction:

$$\sum_x Pr(X = x)(1 - (x - m)^2)$$

- Is there an MPL for the mean?

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
$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	X% chance of \$8	or	m% chance of \$8
m+1	X% chance of \$8	or	m+1% chance of \$8
$\vdots$	$\vdots$	$\vdots$	$\vdots$
99	X% chance of \$8	or	99% chance of \$8
100	X% chance of \$8	or	100% chance of \$8

Requires S-O Reduction: “X% chance” $\sim$ “$E[X]$% chance”

Eliciting the Median

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

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

- Is there an MPL?

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
$\vdots$	$\vdots$	$\vdots$	$\vdots$
m	\$8 if $X \geq m$	or	50% chance of \$8
m+1	\$8 if $X \geq m+1$	or	50% chance of \$8
$\vdots$	$\vdots$	$\vdots$	$\vdots$
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 S-O Reduction

Easily altered to elicit any quantile

Summary

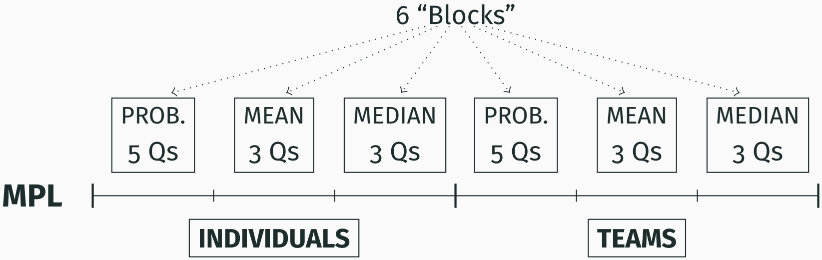
- Six scoring rules:

  Probability: BQSR vs. MPL
  Mean: BQSR vs. MPL
  Median: BLSR vs. MPL

- MPL: weaker assumption for IC (except for the mean)
- MPLs are equiv. to certain scoring rules
- Absolute incentives can be equalized for any p

Experimental Design

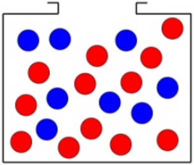
Experimental Design



- Each block has 3 or 5 questions of the same type
- Instructions before each block
- INDIV blocks always precede TEAM blocks
- Order of blocks randomized within INDIV and TEAM
- Order of questions randomized within each block
- Three mechanisms: **MPL**, **BSR**, **NoInfo**
  - Each subject sees only one mechanism

The 11 Questions

This jar contains red and blue marbles.



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

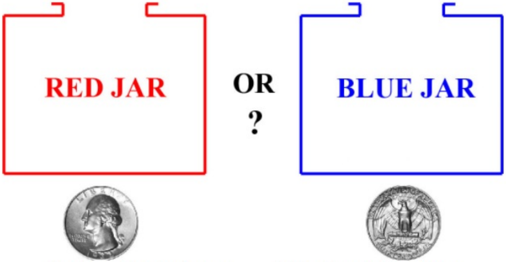
Q1: How many RED marbles are there in the jar? ($ if correct)

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:

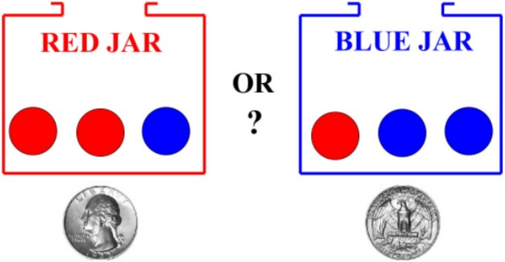


Heads: red jar is chosen. Tails: blue jar is chosen.

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:



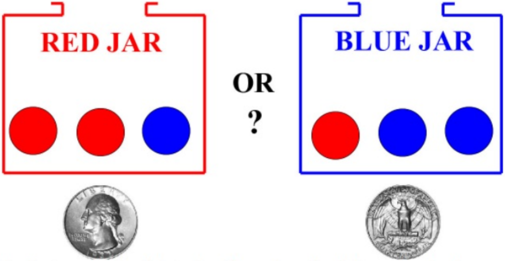
To give you a clue of which jar was chosen, we drew a marble from the chosen jar.

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? ____ %

The 11 Questions

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? ____ %

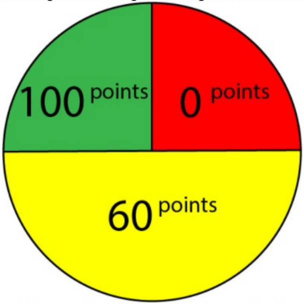
The 11 Questions

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:



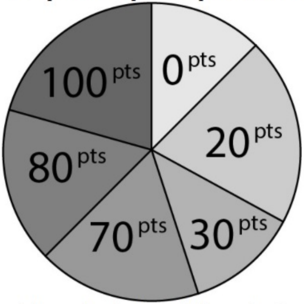
The *median* is the 'middle number.'

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 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 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?

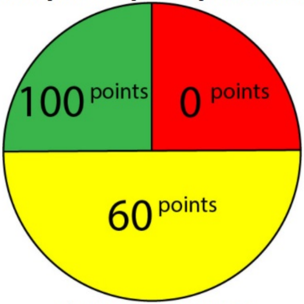
Each question was worth 10 points, for a total of 100.
The *median* is the 'middle number.'
If the median is M, then you have ≥50% chance of getting ≥M points, *and* ≥50% chance of getting ≤M poi

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:

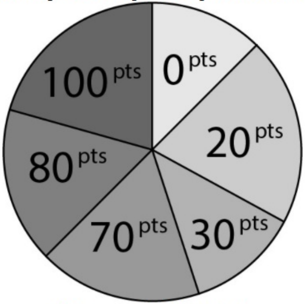


The *mean* is the 'avearge.'
If you multiply each number by its probability and add them up, you get the mean.

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

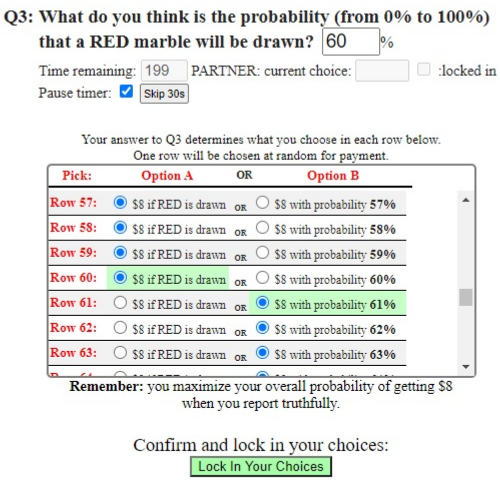
How To Present the Mechanisms

"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)

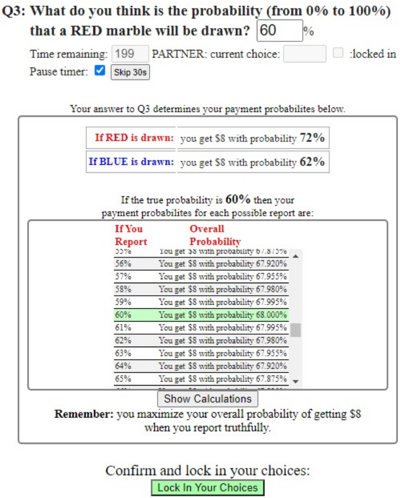
- **BSR:** Wilson & Vespa (2019), Danz, Wilson & Vesterlund (2020)
- **MPL:** Holt & Smith (2016), Healy (2018)

The Mechanism Interfaces: MPL



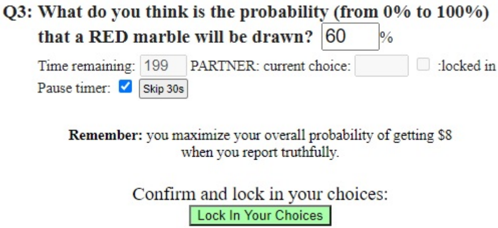
Link

The Mechanism Interfaces: BSR



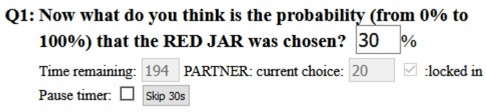
Link

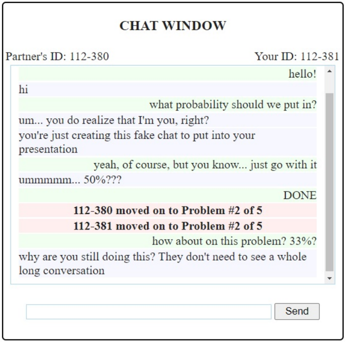
The Mechanism Interfaces: NoInfo



Link

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

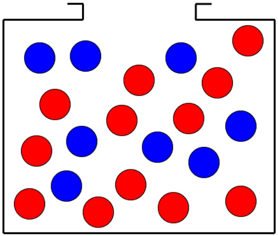
Logistics

- Usual OSU subject pool
- Zoom meeting
- Less control of software environment $\Rightarrow$ missing observations
  - INDIV: 0.7–2.0% TEAM: 4.7–9.3%
- Venmo payments (option for in-person)
- $12 show-up + possible $8 "bonus." (66% won the bonus)
- Still collecting data....

Mechanism:	**MPL**	**BSR**	**NoInfo**
# Subjects:	52	52	47

Results

Objective-Easy #1: % Correct



Pr(Red) = 12/20 = 60%

% Correct:

	MPL	BSR	NoInfo
INDIV:	90.2%	98.1%	95.7%
TEAM:	92.0%	100%	100%

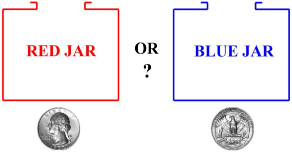
MPL seems worse. Are they trying to manipulate?

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:	89.8%	76.9%	97.9%
TEAM:	100%	92.3%	100%

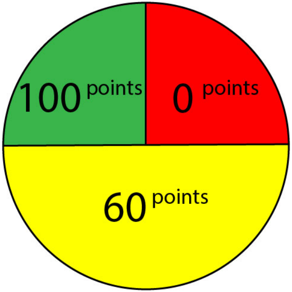
Now BSR seems worse...

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:	74.0%	76.9%	78.7%
TEAM:	81.3%	84.6%	95.2%

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%

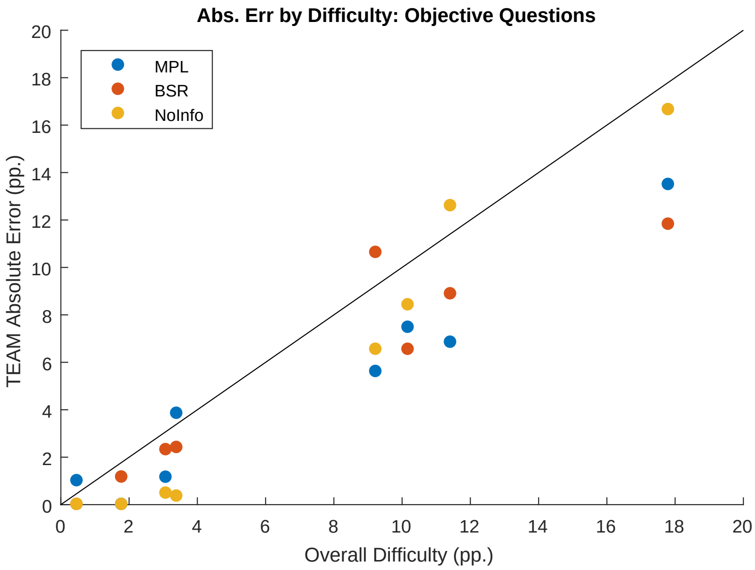
ID#352	MPL	ID#353
		I did 60
55		
		55 is good
55%		55%

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:

Deviate Deviate From Belief

1. May not specify *why* they're deviating

Manipulate Attempt to Manipulate the Payoffs

- May not end up deviating from their belief

Warning: So far, only encoded by me

Chat Encoding

Two Types of Evidence of IC Failures:

Deviate Deviate From Belief

1. May not specify *why* they're deviating

Manipulate Attempt to Manipulate the Payoffs

- May not end up deviating from their belief

Mechanism	MPL	BSR	NoInfo
Deviate	2/26	1/26	0/23
Manipulate	1/26	4/26	0/23

Chat Encoding

Two Types of Evidence of IC Failures:

Deviate Deviate From Belief

1. May not specify *why* they're deviating

Manipulate Attempt to Manipulate the Payoffs

- May not end up deviating from their belief

Mechanism	MPL	BSR	NoInfo
Deviate	2/26	1/26	0/23
Manipulate	1/26	4/26	0/23

Deviations: 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%**

Deviations: BSR

Mean of Hard Quiz Score

ID#305	BSR	ID#306
i have no idea for this one		
		i was just about to say that
but i think 50 gives us the best shot just being right in the middle		
		works for me
50		**50**

Chat Encoding

Two Types of Evidence of IC Failures:

Deviate Deviate From Belief

1. May not specify *why* they're deviating

Manipulate Attempt to Manipulate the Payoffs

- May not end up deviating from their belief

Mechanism	MPL	BSR	NoInfo
Deviate	2/26	1/26	0/23
Manipulate	1/26	4/26	0/23

Manipulations: 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%**

Manipulations: 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**

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

Manipulations: BSR

Mean of Hard Quiz Score

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

(????????)

Manipulations: BSR

Capital of Australia

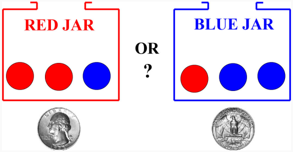
ID#359	BSR	ID#362
this was one i wasnt sure		
i originally thought a high number		
i put 90% but idk		
		i did 48 last time but we can jack up one of the probabilities
		id do 90
Isnt it Syndey? that is pretty well known right?		
		because it gives us 55% chance of getting red and yes it is sydney
	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

(90% $\Rightarrow$ Right: 55%, Wrong: 15%)

The Story

- NoInfo performs the best 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

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%

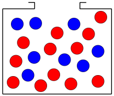
Does The Truth Win?

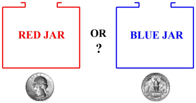
"Truth-Wins" Norm:

2 Right: Both players were correct in INDIV

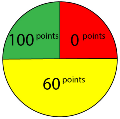
1 Right: One player was correct in INDIV

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



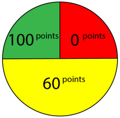


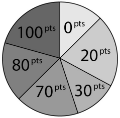
Median

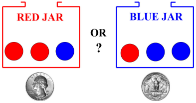


			Median
Won|2 Right:	63/65	52/55	38/43
Won|1 Right:	6/7	15/17	20/25
Won|0 Right:	0/1	1/1	1/5

Does The Truth Win?







	Mean	Median	1 BLUE
Won|2 Right:	24/27	14/18	5/6
Won|1 Right:	20/30	17/31	18/36
Won|0 Right:	4/16	8/24	3/31

Discussion

Summary

- Theory:
    1. MPL has superior IC properties
    2. Some scoring rules are equiv. to an MPL, but not BQSR
- Empirics:
    1. MPL and BSR perform similarly
    2. NoInfo is better when easy, not when hard
    3. Very little evidence of manipulation
        - Subjects are confused/overwhelmed, not manipulating

Recommendations

1. Either mechanism is *fine*
2. Overwhelming details might lead to more mistakes when easy
3. Details might improve belief-formation/calculation when hard

To Do...

1. More observations!!
2. TEAMS first (do they try to manip early?)
   - Can look at errors in "earlier" problems in INDIV
3. More analyses:
   3.1 Encoding confusion/mistakes
   3.2 More analyses of subjective questions
   3.3 Decision time
   3.4 Other suggestions???

Fin