PLEASE STAND BY
Testing Elicitation Mechanisms via Team Chat

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

MiddExLab
Beliefs are central to most economic theories & predictions
Beliefs are central to most economic theories & predictions

Therefore, it’s important that we’re able to measure them accurately
But how should we elicit them?

- **Unincentivized**
  - No: Ramsey (1931), de Finetti (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
Our focus: **BSR** and **MPL**
Our Motivations

• Offerman & Sonnemans (2004): QSR ≃ None
• Armentier & Triech (2013): QSR ⊳ None
• Huck & Weizsacker (2002): QSR ⊳ BDM
• Hollars et al. (2010): BDM ⊳ QSR
• Hao & Houser (2012): BDM-Clock ⊳ BDM
• Hossain & Okui (2013): BSR ⊳ QSR
• Harrison et al. (2014): BSR ≃ QSR-Corr ≳ QSR
• Wilson & Vespa (2017): BSR ⊳ PU-BSR
• Holt & Smith (2016); MPL ⊳ 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. ∃ 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
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
Objective-Easy misreport %’s:

- information ⇒ manipulation!
- Are they really trying to manipulate, or are they just confused?
Our Project

- Have subjects in teams of two, working together via chat
- 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
Leonard J. Savage

THE FOUNDATIONS OF STATISTICS

LEONARD J. SAVAGE

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. 337]:

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. Consequently, some will find parts of it mathematically too...
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 $q = Pr(X = 1)$.

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

$S(q, 1) = 1 - (1 - q)^2$
Two states: $X \in \{0, 1\}$. Announce $q = \Pr(X = 1)$.

For now, assume \textbf{risk neutrality}.
Scoring Rules (Savage 1971)

$\$84 = S(0.4, 0)$
$\$64 = S(0.6, 0)$

$E[S(0.4, X)] = \$76$

Pay if $X = 0$
Pay if $X = 1$

Truthful announcement $\uparrow E[\text{payment}]$
For now, assume risk neutrality
Scoring Rules (Savage 1971)

Pay if $X = 0$

Pay if $X = 1$

Any deviation $\downarrow$ $E[\text{payment}]$

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

Pay if $X = 0$
- $\$50$ at $0\%$
- $\$42$ at $5\%$
- $\$32$ at $10\%$

Pay if $X = 1$
- $\$100$ at $50\%$
- $\$92$ at $60\%$
- $\$82$ at $70\%$

A “flat-to-steep” scoring rule
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 $p$
   • “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

100% = S(0, 0)

84% = S(0.4, 0)

64% = S(0.6, 0)

0% = S(1, 0)

S(1, 1) = 100%;

S(0.6, 1) = 84%;

S(0.4, 1) = 64%;

G(p)
Proof of Incentive Compatibility:

\[
\begin{align*}
    p \cdot S(p, 1) + (1 - p) \cdot S(p, 0) &> p \cdot S(q, 1) + (1 - p) \cdot S(q, 0)
\end{align*}
\]

This requires “Subjective-Objective Reduction”

- Weakening of ROCL
  - Applies only to binary lotteries
- Rules out perceived correlation, probability weighting, etc.
### Multiple Price Lists (MPL)

<table>
<thead>
<tr>
<th>Row#</th>
<th>Option A</th>
<th>OR</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob 1%</td>
</tr>
<tr>
<td>2</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob 2%</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td>$q$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob $q%$</td>
</tr>
<tr>
<td>$q+1$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob $q + 1%$</td>
</tr>
<tr>
<td>$q+2$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob $q + 2%$</td>
</tr>
<tr>
<td>$q+3$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob $q + 3%$</td>
</tr>
<tr>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
<td>⋮</td>
</tr>
<tr>
<td>99</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob 99%$</td>
</tr>
<tr>
<td>100</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 w/ prob 100%$</td>
</tr>
</tbody>
</table>

Choose Option A or Option B (single switch point $q$)
One row randomly selected for payment
<table>
<thead>
<tr>
<th>Row#</th>
<th>Option A</th>
<th>OR</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob 1%</td>
</tr>
<tr>
<td>2</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob 2%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>..</td>
<td>...</td>
</tr>
<tr>
<td>q</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob q%</td>
</tr>
<tr>
<td>q + 1</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob q + 1%</td>
</tr>
<tr>
<td>q + 2</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob q + 2%</td>
</tr>
<tr>
<td>q + 3</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob q + 3%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td>..</td>
<td>...</td>
</tr>
<tr>
<td>99</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob 99%</td>
</tr>
<tr>
<td>100</td>
<td>$8 if X = 1</td>
<td>or</td>
<td>$8 w/ prob 100%</td>
</tr>
</tbody>
</table>

“Multiple Price List” (MPL) version of BDM for probabilities
Holt & Smith (2016), Healy (2018)
### Multiple Price Lists (MPL)

<table>
<thead>
<tr>
<th>Row#</th>
<th>Option A</th>
<th>OR</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } 1%$</td>
</tr>
<tr>
<td>2</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } 2%$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } q%$</td>
</tr>
<tr>
<td>$q+1$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } q+1%$</td>
</tr>
<tr>
<td>$q+2$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } q+2%$</td>
</tr>
<tr>
<td>$q+3$</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } q+3%$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } 99%$</td>
</tr>
<tr>
<td>100</td>
<td>$8 if $X = 1$</td>
<td>or</td>
<td>$8 \text{ w/ prob } 100%$</td>
</tr>
</tbody>
</table>

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.

\[ \iff \]

Subjective-Objective Reduction

\[ \Rightarrow \]

Statewise Dominance

\[ \Rightarrow \]

MPL is I.C.
If you reduce objective lotteries in an MPL, you get a scoring rule
If you reduce objective lotteries in an MPL, you get a scoring rule

Different row probabilities $\Rightarrow$ different $G(p)$
If you reduce objective lotteries in an MPL, you get a scoring rule. Different row probabilities ⇒ 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. $G(1) = 1$
How to equalize incentives across scoring rules?

\[ e.g. \text{ suppose we know } p = 0.3 \]
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 \textbf{mean} for $X$
  • BQSR: $S(m, x) = (1 - (x - m)^2)$
  • 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 $X$

<table>
<thead>
<tr>
<th>Row#</th>
<th>Option A</th>
<th>OR</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$X%$ chance of $$8$</td>
<td>or</td>
<td>$1%$ chance of $$8$</td>
</tr>
<tr>
<td>2</td>
<td>$X%$ chance of $$8$</td>
<td>or</td>
<td>$2%$ chance of $$8$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>$m$</td>
<td>$X%$ chance of $$8$</td>
<td>or</td>
<td>$m%$ chance of $$8$</td>
</tr>
<tr>
<td>$m+1$</td>
<td>$X%$ chance of $$8$</td>
<td>or</td>
<td>$(m+1)%$ chance of $$8$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>...</td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>$X%$ chance of $$8$</td>
<td>or</td>
<td>$99%$ chance of $$8$</td>
</tr>
<tr>
<td>100</td>
<td>$X%$ chance of $$8$</td>
<td>or</td>
<td>$100%$ chance of $$8$</td>
</tr>
</tbody>
</table>

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 $X$

<table>
<thead>
<tr>
<th>Row#</th>
<th>Option A</th>
<th>OR</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$8$ if $X \geq 1$</td>
<td>or</td>
<td>50% chance of $8$</td>
</tr>
<tr>
<td>2</td>
<td>$8$ if $X \geq 2$</td>
<td>or</td>
<td>50% chance of $8$</td>
</tr>
<tr>
<td>$m$</td>
<td>$8$ if $X \geq m$</td>
<td>or</td>
<td>50% chance of $8$</td>
</tr>
<tr>
<td>$m+1$</td>
<td>$8$ if $X \geq m+1$</td>
<td>or</td>
<td>50% chance of $8$</td>
</tr>
<tr>
<td>99</td>
<td>$8$ if $X \geq 99$</td>
<td>or</td>
<td>50% chance of $8$</td>
</tr>
<tr>
<td>100</td>
<td>$8$ if $X \geq 100$</td>
<td>or</td>
<td>50% chance of $8$</td>
</tr>
</tbody>
</table>

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

- **RED JAR**
- **BLUE JAR**

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:

RED JAR

OR

BLUE JAR

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:

[Image of two jars, one red and one blue, each containing red and blue marbles.]

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:

- 100 points
- 0 points
- 60 points

The median is the 'middle number.' If the median is M, then you have \( \geq 50\% \) chance of getting \( \geq M \) points, and \( \geq 50\% \) chance of getting \( \leq M \) 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 \( \geq 50\% \) chance of getting \( \geq M \) points, and \( \geq 50\% \) chance of getting \( \leq M \) points.

Q1: I think the median # of points for this spinner is \( \underline{\hspace{1cm}} \) 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 bibly, 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 \( \geq 50\% \) chance of getting \( \geq M \) points, and \( \geq 50\% \) chance of getting \( \leq M \) points.

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

\[ \underline{\hspace{2cm}} \] pts
The 11 Questions

The computer will spin this spinner one time:

100 points

0 points

60 points

The mean is the 'average.'
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 'average.' 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
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 'average.'

If you multiply each possible score by the probability they got that score and add them up, you get the mean.

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)
**Q3:** What do you think is the probability (from 0% to 100%) that a RED marble will be drawn? 60%

Time remaining: 199
PARTNER: current choice: 
Pause timer: 

Your answer to Q3 determines what you choose in each row below. One row will be chosen at random for payment.

<table>
<thead>
<tr>
<th>Pick:</th>
<th>Option A</th>
<th>OR</th>
<th>Option B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 57:</td>
<td>$8 if RED is drawn</td>
<td>OR</td>
<td>$8 with probability 57%</td>
</tr>
<tr>
<td>Row 58:</td>
<td>$8 if RED is drawn</td>
<td>OR</td>
<td>$8 with probability 58%</td>
</tr>
<tr>
<td>Row 59:</td>
<td>$8 if RED is drawn</td>
<td>OR</td>
<td>$8 with probability 59%</td>
</tr>
<tr>
<td>Row 60:</td>
<td>$8 if RED is drawn</td>
<td>OR</td>
<td>$8 with probability 60%</td>
</tr>
<tr>
<td>Row 61:</td>
<td>$8 with probability 61%</td>
<td>OR</td>
<td></td>
</tr>
<tr>
<td>Row 62:</td>
<td>$8 if RED is drawn</td>
<td>OR</td>
<td>$8 with probability 62%</td>
</tr>
<tr>
<td>Row 63:</td>
<td>$8 with probability 63%</td>
<td>OR</td>
<td></td>
</tr>
</tbody>
</table>

**Remember:** you maximize your overall probability of getting $8 when you report truthfully.

Confirm and lock in your choices:

[Lock In Your Choices]
Q3: What do you think is the probability (from 0% to 100%) that a RED marble will be drawn? 60%.

Time remaining: 169
PARTNER: current choice:  
Pause timer:  
checked

Your answer to Q3 determines your payment probabilities below:

**If RED is drawn:** you get $8 with probability 72%
**If BLUE is drawn:** you get $8 with probability 62%

If the true probability is 60% then your payment probabilities for each possible report are:

<table>
<thead>
<tr>
<th>Overall Probability</th>
<th>Overall Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>60%</td>
<td>You get $8 with probability 68.0000%</td>
</tr>
<tr>
<td>61%</td>
<td>You get $8 with probability 67.9559%</td>
</tr>
<tr>
<td>62%</td>
<td>You get $8 with probability 67.9000%</td>
</tr>
<tr>
<td>63%</td>
<td>You get $8 with probability 67.9559%</td>
</tr>
<tr>
<td>64%</td>
<td>You get $8 with probability 67.9000%</td>
</tr>
<tr>
<td>65%</td>
<td>You get $8 with probability 67.8559%</td>
</tr>
</tbody>
</table>

Show Calculations

Remember: you maximize your overall probability of getting $8 when you report truthfully.

Confirm and lock in your choices:

Lock in Your Choices

Link
Q3: What do you think is the probability (from 0% to 100%) that a RED marble will be drawn? 60% 

Remember: you maximize your overall probability of getting $8 when you report truthfully.

Confirm and lock in your choices:

[Lock In Your Choices]

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

Q1: Now what do you think is the probability (from 0% to 100%) that the RED JAR was chosen? 30%

Time remaining: 194
PARTNER: current choice: 20
Pause timer: □ Skip 30s
• Usual OSU subject pool
• Zoom meeting
• Less control of software environment ⇒ 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....

<table>
<thead>
<tr>
<th>Mechanism:</th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td># Subjects:</td>
<td>52</td>
<td>52</td>
<td>47</td>
</tr>
</tbody>
</table>
Results
Objective-Easy #1: % Correct

Pr(\text{Red}) = \frac{12}{20} = 60\%

% Correct:

<table>
<thead>
<tr>
<th></th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDIV</td>
<td>90.2%</td>
<td>98.1%</td>
<td>95.7%</td>
</tr>
<tr>
<td>TEAM</td>
<td>92.0%</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

MPL seems worse. Are they trying to manipulate?
### Objective-Easy #1: Chats

<table>
<thead>
<tr>
<th>ID#181</th>
<th>MPL</th>
<th>ID#187</th>
</tr>
</thead>
<tbody>
<tr>
<td>i have 12 for red and 8 for blue</td>
<td>12, 20, and 75%? yes</td>
<td>75 sounds good with me</td>
</tr>
<tr>
<td>12</td>
<td>20</td>
<td>75%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID#289</th>
<th>MPL</th>
<th>ID#295</th>
</tr>
</thead>
<tbody>
<tr>
<td>sorry I put wrong answer for 3</td>
<td>12</td>
<td>20</td>
</tr>
</tbody>
</table>
Objective-Easy #2: % Correct

RED JAR OR BLUE JAR

Pr(\text{Red}) = 50\%

% Correct:

<table>
<thead>
<tr>
<th></th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDIV:</td>
<td>89.8%</td>
<td>76.9%</td>
<td>97.9%</td>
</tr>
<tr>
<td>TEAM:</td>
<td>100%</td>
<td>92.3%</td>
<td>100%</td>
</tr>
</tbody>
</table>

Now BSR seems worse...
<table>
<thead>
<tr>
<th>ID#257</th>
<th>BSR</th>
<th>ID#260</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>50% ?</td>
<td></td>
</tr>
<tr>
<td>id say 60</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Why</td>
<td></td>
</tr>
<tr>
<td>cause heads is always more likely</td>
<td>Thats just false</td>
<td></td>
</tr>
<tr>
<td>55 is a compromise</td>
<td>Which is also wrong but whatever</td>
<td></td>
</tr>
<tr>
<td>55%</td>
<td></td>
<td>55%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID#357</th>
<th>BSR</th>
<th>ID#365</th>
</tr>
</thead>
<tbody>
<tr>
<td>(no chat)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>75%</td>
<td></td>
<td>75%</td>
</tr>
</tbody>
</table>
Objective-Easy #3: % Correct

% Correct:

<table>
<thead>
<tr>
<th></th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>INDIV</td>
<td>74.0%</td>
<td>76.9%</td>
<td>78.7%</td>
</tr>
<tr>
<td>TEAM</td>
<td>81.3%</td>
<td>84.6%</td>
<td>95.2%</td>
</tr>
</tbody>
</table>

Median = 60pts
well if it was 100, 0 and 50 the median would be 50 but its 60 and so id go w like 55?

<table>
<thead>
<tr>
<th>ID#343</th>
<th>MPL</th>
<th>ID#345</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>four.osf/five.osf</td>
</tr>
<tr>
<td></td>
<td></td>
<td>five.osf/five.osf/percent.osf</td>
</tr>
<tr>
<td></td>
<td>55%</td>
<td>55%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID#352</th>
<th>MPL</th>
<th>ID#353</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I did 60</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55 is good</td>
<td></td>
</tr>
<tr>
<td></td>
<td>55%</td>
<td>55%</td>
</tr>
<tr>
<td>ID#197</td>
<td>BSR</td>
<td>ID#202</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
<td>--------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>what do u think</td>
</tr>
<tr>
<td></td>
<td></td>
<td>hmm i don’t remember what i said but maybe like 75?</td>
</tr>
<tr>
<td></td>
<td></td>
<td>i’m not sure at all</td>
</tr>
<tr>
<td></td>
<td></td>
<td>love it</td>
</tr>
<tr>
<td></td>
<td>75%</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>75%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ID#302</th>
<th>BSR</th>
<th>ID#308</th>
</tr>
</thead>
<tbody>
<tr>
<td>80?</td>
<td></td>
<td>yeah</td>
</tr>
<tr>
<td>80%</td>
<td></td>
<td>80%</td>
</tr>
</tbody>
</table>
Absolute Error by Treatment

Abs. Err by Difficulty: Objective Questions

- MPL
- BSR
- NoInfo

Overall Difficulty (pp.)

TEAM Absolute Error (pp.)
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
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

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviate</td>
<td>2/26</td>
<td>1/26</td>
<td>0/23</td>
</tr>
<tr>
<td>Manipulate</td>
<td>1/26</td>
<td>4/26</td>
<td>0/23</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviate</td>
<td>2/26</td>
<td>1/26</td>
<td>0/23</td>
</tr>
<tr>
<td>Manipulate</td>
<td>1/26</td>
<td>4/26</td>
<td>0/23</td>
</tr>
</tbody>
</table>
Deviations: MPL

\[
\frac{12}{20} = 60\%
\]

<table>
<thead>
<tr>
<th>ID#352</th>
<th>MPL</th>
<th>ID#353</th>
</tr>
</thead>
<tbody>
<tr>
<td>12 red marbles, 20 total, so 60%</td>
<td></td>
<td>60%</td>
</tr>
<tr>
<td>Yea but I am thinking should we really put the correct number for probability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I mean yeah i think</td>
<td></td>
<td>alright</td>
</tr>
<tr>
<td>Although its random, its the best “odds” then</td>
<td></td>
<td></td>
</tr>
<tr>
<td>60%</td>
<td></td>
<td>60%</td>
</tr>
</tbody>
</table>
## Deviations: BSR

### Mean of Hard Quiz Score

<table>
<thead>
<tr>
<th>ID#305</th>
<th>BSR</th>
<th>ID#306</th>
</tr>
</thead>
<tbody>
<tr>
<td>i have no idea for this one</td>
<td></td>
<td>i was just about to say that</td>
</tr>
<tr>
<td>but i think 50 gives us the best shot</td>
<td></td>
<td>just being right in the middle</td>
</tr>
<tr>
<td>just being right in the middle</td>
<td></td>
<td>works for me</td>
</tr>
</tbody>
</table>

| 50 |  | 50 |
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

<table>
<thead>
<tr>
<th>Mechanism</th>
<th>MPL</th>
<th>BSR</th>
<th>NoInfo</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviate</td>
<td>2/26</td>
<td>1/26</td>
<td>0/23</td>
</tr>
<tr>
<td>Manipulate</td>
<td>1/26</td>
<td>4/26</td>
<td>0/23</td>
</tr>
</tbody>
</table>
12/20/60%

<table>
<thead>
<tr>
<th>ID#352</th>
<th>MPL</th>
<th>ID#353</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>60%</td>
<td></td>
</tr>
</tbody>
</table>

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% | 60% |
### Manipulations: BSR

**Mean of Hard Quiz Score**

<table>
<thead>
<tr>
<th>ID#298</th>
<th>BSR</th>
<th>ID#312</th>
</tr>
</thead>
<tbody>
<tr>
<td>it sounds like 50 but if I took this test I might get 3/4 right</td>
<td>it looks like pretty much any number I type in I get 51/5%</td>
<td>its the same no matter what we type is what I've seen</td>
</tr>
<tr>
<td>50</td>
<td>50</td>
<td></td>
</tr>
</tbody>
</table>

\[(X = M \Rightarrow 51.5\%\)
Manipulations: BSR

<table>
<thead>
<tr>
<th>ID#299</th>
<th>BSR</th>
<th>ID#303</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>40</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>ok</td>
<td>40</td>
</tr>
</tbody>
</table>

40 technically gives the best odds

ok

(?????????)
Capital of Australia

<table>
<thead>
<tr>
<th>ID#359</th>
<th>BSR</th>
<th>ID#362</th>
</tr>
</thead>
<tbody>
<tr>
<td>this was one I wasn't sure, I originally thought a high number, I put 90% but IDK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I did 48 last time but we can jack up one of the probabilities, I'd do 90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Isn't it Sydney? That is pretty well known right?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Because it gives us 55% chance of getting red and yes it is Sydney, everyone knows that because of Finding Nemo lol</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| 90 | 90 |

(90% ⇒ Right: 55%, Wrong: 15%)
• 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, ↑ mistakes
  - Harder problems: details maybe help focus, ↓ mistakes
Errors in Bayesian Updating

- One Blue Draw:
  - \( Pr(R|b) = Pr(R) \times Pr(b|R) \). 17%
  - Marble draw is uninformative. 50%

- Two Blue Draws:
  - \( Pr(R|bb) = Pr(R) \times Pr(b|R) \times Pr(b|R) \). 6%
  - Second draw gives no new info. Same as one.
  - Marble draws are uninformative. 50%
  - Second draw was with replacement. 0%
“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)

| Won|2 Right| Won|1 Right| Won|0 Right|
|---|---|---|---|---|
| 63/65 | 52/55 | 38/43 |
| 6/7 | 15/17 | 20/25 |
| 0/1 | 1/1 | 1/5 |
Does The Truth Win?

**Won|2 Right:** 24/27  
**Won|1 Right:** 20/30  
**Won|0 Right:** 4/16

**Mean**
- 100 points
- 60 points

**Median**
- 100 pts
- 80 pts
- 70 pts
- 30 pts
- 20 pts
- 0 pts

**1 BLUE**

**RED JAR**
- 5/6

**BLUE JAR**
- 18/36

- 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
1. Either mechanism is *fine*

2. Overwhelming details might lead to more mistakes when easy

3. Details might improve belief-formation/calculation when hard
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