Robust Bayesian Truth Serum

Presentation by Mark Bun and Bo Waggoner

2012-12-03
Outline

1. Introduction and Setting
   - Recap: Human Computation Mechanisms so far
   - The RBTS Approach
   - Setting

2. RBTS and Shadowing
   - Shadowing
   - RBTS

3. PP Without Common Prior

4. Summary: Human Computation Mechanisms
   - Assumptions and Results
Recap

Peer prediction:
- Elicit?
- Rewards?
Recap

Peer prediction:
- Elicit?
- Rewards?

Bayesian Truth Serum:
- Elicit?
- Rewards?
Big problem with implementing PP?
Big problem with implementing PP?

- Mechanism needs to know the information structure!
Big problem with implementing PP?
- Mechanism needs to know the information structure!

Big problem with BTS?
Big problem with implementing PP?
- Mechanism needs to know the information structure!

Big problem with BTS?
- Requires $n \to \infty$!
Big problem with implementing PP?
- Mechanism needs to know the information structure!

Big problem with BTS?
- Requires $n \to \infty$!

Other problems with BTS? Does RBTS resolve them?
Goal: Truthful mechanism for any $n$ that doesn’t rely on the mechanism knowing the information structure.
Goal: Truthful mechanism for any $n$ that doesn’t rely on the mechanism knowing the information structure.

Approach:

- Recall: Elicit two-part report (prediction and signal)
- Which one is easy to incentivize and which is difficult?
Goal: Truthful mechanism for any $n$ that doesn’t rely on the mechanism knowing the information structure.

Approach:
- Recall: Elicit two-part report (prediction and signal)
- Which one is easy to incentivize and which is difficult?

To elicit the signal: Use shadowing!
Belief system

Belief system consists of state $T$ drawn from $\{1, \ldots, m\}$ and signals $S$ drawn from $\{0, 1\}$. Agents have common prior $Pr[T = t]$ and $Pr[S = h | T = t]$.

"Impersonally informative": For every $i, j, k$, write $p\{s_i\} = Pr[S_j = h | S_i = s_i]$ and $p\{s_i, s_j\} = Pr[S_k = h | S_i = s_i, S_j = s_j]$. 

Presentation by Mark Bun and Bo Waggoner

Robust Bayesian Truth Serum
Belief system

- Consists of state $T$ drawn from $\{1, \ldots, m\}$ and signals $S$ drawn from $\{l, h\} = \{0, 1\}$
Belief system

- Consists of state $T$ drawn from $\{1, \ldots, m\}$ and signals $S$ drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$. 
Belief system

- Consists of state $T$ drawn from $\{1, \ldots, m\}$ and signals $S$ drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.
- “Impersonally informative”
Belief system

- Consists of state $T$ drawn from $\{1, \ldots, m\}$ and signals $S$ drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.
- “Impersonally informative”: For every $i, j, k$, write

$$p_{\{s_i\}} = \Pr[S_j = h | S_i = s_i]$$
Belief system

- Consists of state $T$ drawn from $\{1, \ldots, m\}$ and signals $S$ drawn from $\{l, h\} = \{0, 1\}$
- Agents have common prior $\Pr[T = t]$ and $\Pr[S = h | T = t]$.
- “Impersonally informative” : For every $i, j, k$, write

$$p_{\{s_i\}} = \Pr[S_j = h | S_i = s_i]$$
$$p_{\{s_i, s_j\}} = \Pr[S_k = h | S_i = s_i, S_j = s_j]$$
Admissibility

1. Admissible prior: at least two states ($m \geq 2$)
2. Every state has positive probability ($\text{Pr}[T = t] > 0$)
3. Assortative property:
   \[
   \text{Pr}[S = h | T = 1] < \cdots < \text{Pr}[S = h | T = m]
   \]
4. Fully mixed: $0 < \text{Pr}[S = h | T = t] < 1$

Lemma 6: For an admissible prior,

\[
\frac{\text{Pr}\{h, h\}}{\text{Pr}\{h\}} > \frac{\text{Pr}\{h, l\}}{\text{Pr}\{l\}} > \frac{\text{Pr}\{l, h\}}{\text{Pr}\{l\}} > 0
\]
Admissibility

Admissible prior:
Introduction and Setting
RBTS and Shadowing
PP Without Common Prior
Summary: Human Computation Mechanisms

Recap: Human Computation Mechanisms so far
The RBTS Approach
Setting

Admissibility

Admissible prior:
1. At least two states \( m \geq 2 \)
Admissibility

**Admissible prior:**

1. At least two states \( m \geq 2 \)
2. Every state has positive probability \( \Pr[T = t] > 0 \)
Admissibility

Admissible prior:

1. At least two states \((m \geq 2)\)
2. Every state has positive probability \((\Pr[T = t] > 0)\)
3. Assortative property:
   \[
   \Pr[S = h | T = 1] < \cdots < \Pr[S = h | T = m]
   \]
Admissibility

Admissible prior:

1. At least two states \( m \geq 2 \)
2. Every state has positive probability \( \Pr[T = t] > 0 \)
3. Assortative property:
   \[ \Pr[S = h | T = 1] < \cdots < \Pr[S = h | T = m] \]
4. Fully mixed:
   \[ 0 < \Pr[S = h | T = t] < 1 \]
Admissibility

Admissible prior:

1. At least two states ($m \geq 2$)
2. Every state has positive probability ($\Pr[T = t] > 0$)
3. Assortative property:
   \[
   \Pr[S = h | T = 1] < \cdots < \Pr[S = h | T = m]
   \]
4. Fully mixed: $0 < \Pr[S = h | T = t] < 1$

Lemma 6: For an admissible prior,

\[
1 > p\{h,h\} > p\{h\} > p\{h,l\} = p\{l,h\} > p\{l\} > p\{l,l\} > 0
\]
What does shadowing solve?

Truthfully elicit signals instead of beliefs $\omega \in \{0, 1\}$ a binary future event. Agent $i$ draws a signal $S_i \in \{0, 1\} = \{l, h\}$. How do we get agent $i$ to truthfully reveal $S_i$?
What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
- \( \omega \in \{0, 1\} \) a binary future event
What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
- \( \omega \in \{0, 1\} \) a binary future event
- Agent \( i \) draws a signal \( S_i \in \{0, 1\} = \{l, h\} \)
What does shadowing solve?

- Truthfully elicit *signals* instead of *beliefs*
- \( \omega \in \{0, 1\} \) a binary future event
- Agent \( i \) draws a signal \( S_i \in \{0, 1\} = \{l, h\} \)
- How do we get agent \( i \) to truthfully reveal \( S_i \)?
Naïve approach

Choose a strictly proper binary scoring rule $R(y, \omega)$.
Ask agent $i$ for signal report $x_i \in \{0, 1\}$ using $R(x_i, \omega)$.

Might not work if $i$'s beliefs $p_{\{S_i\}}$ about $\omega$ are far from $x_i$!
Naïve approach

1. Choose a strictly proper binary scoring rule $R(y, \omega)$
Naïve approach

1. Choose a strictly proper binary scoring rule $R(y, \omega)$
2. Ask agent $i$ for signal report $x_i \in \{0, 1\}$ using $R(x_i, \omega)$
Naïve approach

1. Choose a strictly proper binary scoring rule $R(y, \omega)$
2. Ask agent $i$ for signal report $x_i \in \{0, 1\}$ using $R(x_i, \omega)$

Might not work if $i$’s beliefs $p_{S_i}$ about $\omega$ are far from $x_i$!
Solution: Use a reference prediction
Solution: Use a reference prediction

Let \( y \in (0, 1) \) be an arbitrary prediction report.
Solution: Use a reference prediction

1. Let $y \in (0, 1)$ be an arbitrary prediction report
2. Set $\delta = \min(y, 1 - y)$
Solution: Use a reference prediction

1. Let \( y \in (0, 1) \) be an arbitrary prediction report.
2. Set \( \delta = \min(y, 1 - y) \).
3. Create a shadow report for \( i \):
   \[
   y'_i = \begin{cases} 
   y + \delta & \text{if } x_i = 1 \\
   y - \delta & \text{if } x_i = 0 
   \end{cases}
   \]
Solution: Use a reference prediction

1. Let $y \in (0, 1)$ be an arbitrary prediction report.
2. Set $\delta = \min(y, 1 - y)$.
3. Create a shadow report for $i$:
   \[ y'_i = \begin{cases} 
   y + \delta & \text{if } x_i = 1 \\
   y - \delta & \text{if } x_i = 0 
   \end{cases} \]
4. Score agent $x_i$ using the quadratic scoring rule $R_q(y'_i, \omega)$. 
When and why does shadowing work?
When and why does shadowing work?

- Intuition: Truthfully reporting \( x_i = S_i \) pulls the reference prediction toward \( i \)'s posterior
When and why does shadowing work?

- Intuition: Truthfully reporting $x_i = S_i$ pulls the reference prediction toward $i$’s posterior.
- Selten ’98: Let $Y \subset [0, 1]$. If agent $i$ has beliefs $p$, then she maximizes her expected score $R_q(y, \omega)$ over $y \in Y$ by minimizing $|y - p|$. 

**Lemma 8**: Suppose agent $i$ has observed signals $I \in \{l, h\}^k$ and $S_i$. If $p\{l, I\} < y < p\{h, I\}$, then agent $i$ should truthfully report $x_i = S_i$. 

Presentation by Mark Bun and Bo Waggoner

Robust Bayesian Truth Serum
When and why does shadowing work?

- **Intuition:** Truthfully reporting $x_i = S_i$ pulls the reference prediction toward $i$’s posterior.

- **Selten ’98:** Let $Y \subset [0, 1]$. If agent $i$ has beliefs $p$, then she maximizes her expected score $R_q(y, \omega)$ over $y \in Y$ by minimizing $|y - p|$.

- **Lemma 8:** Suppose agent $i$ has observed signals $I \in \{l, h\}^k$ and $S_i$. If $p\{l, I\} < y < p\{h, I\}$, then agent $i$ should truthfully report $x_i = S_i$. 
What’s missing?

How do we pick $y$ to guarantee $p\{l, I\} < y < p\{h, I\}$?

How do we pick $\omega$ when there is no ground truth?
What’s missing?

- How do we pick $y$ to guarantee $p_{l,I} < y < p_{h,I}$?
What’s missing?

- How do we pick $y$ to guarantee $p_{\{l,I\}} < y < p_{\{h,I\}}$?
- How do we pick $\omega$ when there is no ground truth?
Robust Bayesian Truth Serum
Robust Bayesian Truth Serum

- Idea: Shadowing + reference raters
Robust Bayesian Truth Serum

- **Idea**: Shadowing + reference raters
- **Information report**: $x_i \in \{0, 1\}$ represents agent $i$’s signal
Robust Bayesian Truth Serum

- Idea: Shadowing + reference raters
- **Information report:** $x_i \in \{0, 1\}$ represents agent $i$’s signal
- **Prediction report:** $y_i \in [0, 1]$ represents agent $i$’s prediction for the frequency of $h$ signals
Scoring

For each agent $i$, pick reference raters $j, k$.
\[ y = y_j \text{ and } \omega = x_k. \]
Create a shadow report for $i$:
\[ y'_i = \begin{cases} 
  y_j + \delta & \text{if } x_i = 1 \\
  y_j - \delta & \text{if } x_i = 0
\end{cases} \]
where\[ \delta = \min(y_j, 1 - y_j). \]

RBTS score:
\[ R_q(y'_i, x_k) + R_q(y_i, x_k). \]
Scoring

- For each agent $i$, pick reference raters $j, k$
Scoring

- For each agent $i$, pick reference raters $j, k$
- Set $y = y_j$ and $\omega = x_k$. 
Scoring

- For each agent $i$, pick reference raters $j, k$
- Set $y = y_j$ and $\omega = x_k$.
- Create a shadow report for $i$:

$$y'_i = \begin{cases} y_j + \delta & \text{if } x_i = 1 \\ y_j - \delta & \text{if } x_i = 0 \end{cases}$$

where $\delta = \min(y_j, 1 - y_j)$
Scoring

- For each agent $i$, pick reference raters $j$, $k$
- Set $y = y_j$ and $\omega = x_k$.
- Create a shadow report for $i$:

\[ y'_i = \begin{cases} 
    y_j + \delta & \text{if } x_i = 1 \\
    y_j - \delta & \text{if } x_i = 0
\end{cases} \]

where $\delta = \min(y_j, 1 - y_j)$

- **RBTS score:**

\[ R_q(y'_i, x_k) + R_q(y_i, x_k) \]

\[ \text{information score} + \text{prediction score} \]
Incentive compatibility
Incentive compatibility

- If \( n \geq 3 \) agents have signals drawn from an admissible common prior, then reporting truthfully is a strict Bayes-Nash equilibrium.
If $n \geq 3$ agents have signals drawn from an \textit{admissible} common prior, then reporting truthfully is a \textit{strict} Bayes-Nash equilibrium.

Why should agents report their predictions truthfully?
If $n \geq 3$ agents have signals drawn from an admissible common prior, then reporting truthfully is a strict Bayes-Nash equilibrium.

Why should agents report their predictions truthfully?

How about signals?
Incentive compatibility

- If \( n \geq 3 \) agents have signals drawn from an **admissible** common prior, then reporting truthfully is a **strict Bayes-Nash equilibrium**.
- Why should agents report their predictions truthfully?
- How about signals? Enough to show that admissibility

\[
p\{i, I\} < y_j < p\{h, I\}.
\]
Incentive compatibility

Lemma 6:

For an admissible prior,

\[ p\{h,h\} > p\{h\} > p\{h,l\} = p\{l,h\} > p\{l\} > p\{l,l\} > 0 \]

Case 1:

\[ y_j = p\{h\} = \Rightarrow p\{l,h\} < y_j = p\{h\} < p\{h,h\} \]

Case 2:

\[ y_j = p\{l\} = \Rightarrow p\{l,l\} < y_j = p\{l\} < p\{h,l\} \]

Recap:

\[ p\{l,S_j\} < y_j < p\{h,S_j\} \]

so agent \(i\) should truthfully reveal \(x_i = S_i\).
Incentive compatibility

**Lemma 6:** For an admissible prior,

\[ 1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0 \]
Incentive compatibility

- **Lemma 6:** For an admissible prior,

\[ 1 > p_{\{h,h\}} > p_{\{h\}} > p_{\{h,l\}} = p_{\{l,h\}} > p_{\{l\}} > p_{\{l,l\}} > 0 \]

- **Case 1:** \( y_j = p_{\{h\}} \implies p_{\{l,h\}} < y_j = p_{\{h\}} < p_{\{h,h\}} \)
**Lemma 6:** For an admissible prior,

\[ 1 > p_{h,h} > p_h > p_{h,l} = p_{l,h} > p_l > p_{l,l} > 0 \]

- **Case 1:** \( y_j = p_h \) \( \implies \) \( p_{l,h} < y_j = p_h < p_{h,h} \)
- **Case 2:** \( y_j = p_l \) \( \implies \) \( p_{l,l} < y_j = p_l < p_{h,l} \)
Incentive compatibility

- **Lemma 6:** For an admissible prior,

\[ 1 > p\{h,h\} > p\{h\} > p\{h,l\} = p\{l,h\} > p\{l\} > p\{l,l\} > 0 \]

- **Case 1:** \( y_j = p\{h\} \implies p\{l,h\} < y_j = p\{h\} < p\{h,h\} \)
- **Case 2:** \( y_j = p\{l\} \implies p\{l,l\} < y_j = p\{l\} < p\{h,l\} \)
- Recap: \( p\{l,S_j\} < y_j < p\{h,S_j\} \), so agent \( i \) should truthfully reveal \( x_i = S_i \)
Properties of RBTS
Properties of RBTS

- Works for any number of agents $n \geq 3$
Properties of RBTS

- Works for any number of agents $n \geq 3$
- Scores are well-defined between 0 and 2.
Properties of RBTS

- Works for any number of agents $n \geq 3$
- Scores are well-defined between 0 and 2. Participation is “ex post individually rational”.
Properties of RBTS

• Works for any number of agents \( n \geq 3 \)
• Scores are well-defined between 0 and 2. Participation is “ex post individually rational”.
• Numerically robust
Extensions
Extensions

- Adapting to a budget constraint.
Extensions

- Adapting to a budget constraint. Scaling, randomized exclusion.
Extensions

- Adapting to a budget constraint. Scaling, randomized exclusion.
- More than two outcomes?
Extensions

- Adapting to a budget constraint. Scaling, randomized exclusion.
- More than two outcomes?
- Other proper scoring rules?

- Same setting as RBTS except each agent may have a different information structure.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent’s prior and posterior probability that another agent will get a high signal.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent’s prior and posterior probability that another agent will get a high signal.
- Infer whether agent’s signal was high or low.
**Peer Prediction Without a Common Prior.** Witkowski, Parkes. EC-2012.

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent’s prior *and* posterior probability that another agent will get a high signal.
- Infer whether agent’s signal was high or low
- Score = score(prior) + score(posterior)

- Same setting as RBTS except each agent may have a different information structure.
- Elicit both the agent’s prior and posterior probability that another agent will get a high signal.
- Infer whether agent’s signal was high or low
- Score $= \text{score(prior)} + \text{score(posterior)}$
- OR elicit prior and signal, and shadow to get a posterior (complicated)
<table>
<thead>
<tr>
<th></th>
<th>PP</th>
<th>PP-CP</th>
<th>BTS</th>
<th>RBTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>common prior exists</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>designer knows CP</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>information structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of players</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># of signals</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>designer learns?</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>