Trick or Treat: Putting Peer Prediction to the Test

Yousef Alhessi
CS 97
Why Peer-prediction Mechanisms?

Q: Can we design mechanisms to truthfully elicit subjective evaluations from individual?

A: Try peer-prediction mechanisms!
How do they work?

● One participant’s report is compared to those of their peers to induce truthful reporting equilibrium.
Example

- Let’s say I make cameras.
- I want to learn about the quality of the cameras from the consumers.
- Ask the consumers to rate the product and compare their ratings.
- Pay them for rating!
How do we do that?

- With proper incentives, it is in a participant’s best interest to report truthfully if they believe all other participants will also be truthful.
But there is a problem...

- There exists at least two other uninformative equilibria.
- Participants can coordinate to give the same answer over and over again.
The trick or treat story
The trick or treat story
The trick or treat story
The trick or treat story

What candy did you get?
Payment Rules

<table>
<thead>
<tr>
<th></th>
<th>ref report = GB</th>
<th>ref report = MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>report = GB</td>
<td>$1.20</td>
<td>$0.30</td>
</tr>
<tr>
<td>report = MM</td>
<td>$0.10</td>
<td>$1.50</td>
</tr>
</tbody>
</table>

Table 1: Typical Payment Rule

<table>
<thead>
<tr>
<th></th>
<th>ref report = GB</th>
<th>ref report = MM</th>
</tr>
</thead>
<tbody>
<tr>
<td>report = GB</td>
<td>$1.50</td>
<td>$0.10</td>
</tr>
<tr>
<td>report = MM</td>
<td>$0.10</td>
<td>$1.50</td>
</tr>
</tbody>
</table>

Table 2: Payment Rule Rewarding Agreement
The trick or treat story

You are playing **20 rounds** of the game with **2 other player(s)**. Your bonus (= average reward) so far: **$0.63**.

Current Status: ***Please confirm your claim!***

Current Round is #4:
1. A house has been randomly chosen for this round.

   - 50% chance of winning 80% of the candy
   - 50% chance of winning 70% of the candy

   The randomly chosen house

2. Get a random candy from the chosen house.

   Your candy is: 🍫

3. Choose your claim.

   ![Confirm your claim!]

   **Reward Rule**

<table>
<thead>
<tr>
<th>Your claim</th>
<th>Other player's claim</th>
<th>Your reward</th>
</tr>
</thead>
<tbody>
<tr>
<td>🍫</td>
<td>🍫</td>
<td>$0.30</td>
</tr>
<tr>
<td>🍫</td>
<td>🍫</td>
<td>$1.50</td>
</tr>
<tr>
<td>🍫</td>
<td>🍫</td>
<td>$0.10</td>
</tr>
<tr>
<td>🍫</td>
<td>🍫</td>
<td>$1.20</td>
</tr>
</tbody>
</table>

   Round | Your candy | Your claim | Other players’ claims | Your reward |
   ------|------------|------------|-----------------------|-------------|
   1     | 🍫         | 🍫         | 🍫 1 1                | $0.10       |
   2     | 🍫         | 🍫         | 🍫 1 1                | $0.30       |
   3     | 🍫         | 🍫         | 🍫 1 1                | $1.50       |
   4     | 🍫         | 🍫         | ❓ 1 of 2 other player(s) confirmed claims | |

   **Average reward: $0.63**

Figure 1: The Game Interface
Results

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>GB</th>
<th>Truthful</th>
<th>Unclassified</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>47</td>
<td>4</td>
<td>5</td>
<td>47</td>
<td>103</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>7</td>
<td>34</td>
<td>7</td>
<td>56</td>
<td>104</td>
</tr>
</tbody>
</table>

Table 3: Simple equilibrium convergence classification

<table>
<thead>
<tr>
<th></th>
<th>MM</th>
<th>GB</th>
<th>Truthful</th>
<th>Unclassified</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>62</td>
<td>11</td>
<td>18</td>
<td>12</td>
<td>103</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>12</td>
<td>47</td>
<td>22</td>
<td>23</td>
<td>104</td>
</tr>
</tbody>
</table>

Table 4: Relaxed equilibrium convergence classification
Relaxed Results

(a) Aggregate strategies in treatment 1.

(b) Aggregate strategies in treatment 2.
Summary

- Just using MRZ as a peer prediction mechanism the way doesn’t work in practice.
Discussion Questions

- Can we make peer prediction mechanisms work somehow?
- Would adding artificial honest players to each game make the truthful equilibrium dominate?
- What about a special payoff matrix?