Epistemic Experiments: Utilities, Beliefs, and Irrational Play

Paul J. Healy (OSU)
An Embarrassing Timeline

A pet project:

2010: Tenure. First attempt at $2 \times 2$ games.

2013: Redo experiment on pencil & paper

2014: Present at Pitt

2015: Add centipede games

2016: Add no-elicitation benchmark

2017: Add sequential-move $2 \times 2$ games

2019: “I’m never presenting this again.”

2020: COVID writing retreat, 1st draft

2023: Present at Pitt
The Standard Approach

Standard game theory experiment:

1. Interesting game form
The Standard Approach

Standard game theory experiment:

1. Interesting game form
2. Baseline theory/assumptions:
   • Selfish prefs, “Rational” behavior (eg, backwards induction)
The Standard Approach

Standard game theory experiment:

1. Interesting game form
2. Baseline theory/assumptions:
   • Selfish prefs, “Rational” behavior (eg, backwards induction)
3. Observe deviations

![Graph showing different outcomes and payoffs in a game theoretical context.](image-url)
The Standard Approach

Standard game theory experiment:

1. Interesting game form
2. Baseline theory/assumptions:
   • Selfish prefs, “Rational” behavior (eg, backwards induction)
3. Observe deviations
4. Posit alternative theory
The Standard Approach

Standard game theory experiment:

1. Interesting game form
2. Baseline theory/assumptions:
   • Selfish prefs, “Rational” behavior (eg, backwards induction)
3. Observe deviations
4. Posit alternative theory
5. New treatments to test comparative statics
Alternative “Solution Concepts”

1. Nash with Altruism, Inequality Aversion
2. Reputation-building/Gang of Four
3. Level-$k$ (wrong beliefs)
4. QRE (noisy equilibrium play)
A Risky Picture

strategies

$S_1$  $S_2$  $S_3$  $S_4$

selfish  altruist  level 1  level 2  level 3  ... beliefs, prefs, etc

correct  correct  belief  belief
Each solution concept makes specific assumptions about utilities, beliefs, rationality, etc.

Why not measure these things directly???
Direct Measurement

Each solution concept makes specific assumptions about utilities, beliefs, rationality, etc.

Why not measure these things directly???

(Yes, eliciting these things might change behavior. I’ll get to that.)
Each solution concept makes specific assumptions about utilities, beliefs, rationality, etc.

Why not measure these things directly???

(Yes, eliciting these things might change behavior. I'll get to that.)

OK... but then what exactly should we measure?
Each solution concept makes specific assumptions about utilities, beliefs, rationality, etc.

Why not measure these things directly???

(Yes, eliciting these things might change behavior. I’ll get to that.)

OK... but then what exactly should we measure?

We need a framework that encompasses all such theories

...a level playing field in which no theory is the null hypothesis
The Epistemic Game Theory Framework

The Observable Experiment: \((I, S, X, \pi)\)

1. Players: \(i \in I = \{1, 2\}\)
2. Strategies: \(s_i \in S_i \quad \text{Ex: when to Take}\)
3. Outcomes: \((x_1, x_2) \in X \quad \text{Ex: ($6.40, $1.60$)}\)
4. Outcome function: \(\pi(s_1, s_2) \in X\)

\(i\)'s Private Information: \(\omega_i = (u_i, s_i, \vec{p}_i)\)

1. Utility: \(u_i(x_1, x_2)\)
   - Non-selfish, but consequentialist
2. Chosen Strategy: \(s_i \in S_i\)
   - Mixing is only in beliefs (Aumann)
3. Beliefs
   - First-order: \(p_i^1(u_{-i}, s_{-i})\)
   - Second-order: \(p_i^2(p_{-i}^1, u_{-i}, s_{-i})\)
   - Hierarchy: \(\vec{p}_i = (p_i^1, p_i^2, p_i^3, \ldots)\)
Players are in a (selfish) Nash equilibrium at $\omega = (\omega_1, \ldots, \omega_n)$ if:

1. Utility: $u_i(x_i, x_{-i}) = x_i$ ("selfish")
2. Beliefs: correct beliefs about $u_{-i}$, $s_{-i}$.
3. Strategy: $s_i \in \arg \max_{s_i} \left[ \sum_{(u_{-i}, s_{-i})} p_i^1(u_{-i}, s_{-i})u_i(\pi(s'_i, s_{-i})) \right]$
Players are in a (selfish) Nash equilibrium at \( \omega = (\omega_1, \ldots, \omega_n) \) if:

1. Utility: \( u_i(x_i, x_{-i}) = x_i \) ("selfish")
2. Beliefs: correct beliefs about \( u_{-i}, s_{-i} \).
3. Strategy: \( s_i \in \arg \max_{s'_i} \left[ \sum_{(u_{-i}, s_{-i})} p^1_i(u_{-i}, s_{-i})u_i(\pi(s'_i, s_{-i})) \right] \)

   - Player \( i \) is rational at \( \omega_i = (u_i, s_i, \hat{p}_i) \) if this is true
   - Let \( R_i \) be those \( (p^1_i, u_i, s_i) \) where \( i \) is rational
   - \( i \)'s belief that \(-i\) is rational is \( p^2_i(R_{-i}) \)
   - Can define common knowledge of rationality, etc.
Players are in a (selfish) Nash equilibrium at $\omega = (\omega_1, \ldots, \omega_n)$ if:

1. Utility: $u_i(x_i, x_{-i}) = x_i$ ("selfish")
2. Beliefs: correct beliefs about $u_{-i}, s_{-i}$.
3. Strategy: $s_i \in \arg \max_{s'_i} \left[ \sum_{(u_{-i}, s_{-i})} p^1_i(u_{-i}, s_{-i})u_i(\pi(s'_i, s_{-i})) \right]$

   - Player $i$ is **rational** at $\omega_i = (u_i, s_i, \bar{p}_i)$ if this is true
   - Let $R_i$ be those $(p^1_i, u_i, s_i)$ where $i$ is rational
   - $i$’s belief that $-i$ is rational is $p^2_i(R_{-i})$
   - Can define common knowledge of rationality, etc.
   - Aumann (1995): Nash equil. does **not** require c.k. of rationality
   - Rationality & c.k. of rationality $\Rightarrow$ IESDS
Example: Altruistic Nash Equilibrium

Players are in an **Altruistic** Nash equilibrium at \( \omega = (\omega_1, \ldots, \omega_n) \) if:

1. **Utility:** \( u_i(x_i, x_{-i}) = x_i + \alpha x_{-i} \)
2. **Beliefs:** correct beliefs about \( u_{-i}, s_{-i} \).
3. **Strategy:** \( s_i \in \arg \max_{s'_i} \left[ \sum_{(u_{-i}, s_{-i})} p^1_i(u_{-i}, s_{-i})u_i(\pi(s'_i, s_{-i})) \right] \)
   - Player \( i \) is **rational** at \( \omega_i = (u_i, s_i, \bar{p}_i) \) if this is true
   - Let \( R_i \) be those \((p^1_i, u_i, s_i)\) where \( i \) is rational
   - \( i \)'s belief that \(-i\) is rational is \( p^2_i(R_{-i}) \)
   - Can define common knowledge of rationality, etc.
   - Aumann (1995): Nash equil. does **not** require c.k. of rationality
   - Rationality & c.k. of rationality \( \Rightarrow \) IESDS
Example: Level-$k$

Level-1:

1. Utility: selfish
2. Beliefs: $u_i$ selfish, $s_i$ uniformly distributed
3. Strategy: $s_i$ is rational, given utility & beliefs

Level-$k > 1$:

1. Utility: selfish
2. Beliefs: $u_i$ selfish, $s_i$ is Level-$k - 1$ strategy
3. Strategy: $s_i$ is rational, given utility & beliefs
So, What Should We Measure?

1. Utility: \( u_i(\pi(s_1, s_2)) \)

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td>$2,2$</td>
<td>$0,3$</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>$3,0$</td>
<td>$1,1$</td>
</tr>
</tbody>
</table>

A Game Form

**A Game**

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td>$4,4$</td>
<td>$0,0$</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>$0,0$</td>
<td>$2,2$</td>
</tr>
</tbody>
</table>

A Game

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>U</strong></td>
<td>$0,0$</td>
<td>$0,0$</td>
</tr>
<tr>
<td><strong>D</strong></td>
<td>$0,0$</td>
<td>$0,0$</td>
</tr>
</tbody>
</table>

A Game
How To Measure Cardinal Utility

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>$2,$2</td>
<td>$0,$3</td>
</tr>
<tr>
<td>D</td>
<td>$3,$0</td>
<td>$1,$1</td>
</tr>
</tbody>
</table>

A Game Form

- Elicit $u_i(x_1, x_2)$ for each cell
  - or, for each terminal node

- How?
  - Let $\bar{x} = ($20, $20), x = ($0, $0)$
  - “I’m indifferent between ($3, $0)$ and getting $\bar{x}$ w/ prob. $p^*$”

$$u_i($3, $0) = p^* \underbrace{u_i(\bar{x})}_{=1} + (1 - p^*) \underbrace{u_i(x)}_{=0} = p^*$$
## Multiple Price List Elicitation

<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>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob 1%</td>
</tr>
<tr>
<td>2</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob 2%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>q</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob q%</td>
</tr>
<tr>
<td>q + 1</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob q + 1%</td>
</tr>
<tr>
<td>q + 2</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob q + 2%</td>
</tr>
<tr>
<td>q + 3</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob q + 3%</td>
</tr>
<tr>
<td>...</td>
<td>...</td>
<td></td>
<td>...</td>
</tr>
<tr>
<td>99</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) w/ prob 99%</td>
</tr>
<tr>
<td>100</td>
<td>($3,$0)</td>
<td>or</td>
<td>($20,$20) 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>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob 1%</td>
</tr>
<tr>
<td>2</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob 2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$q$</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob $q%$</td>
</tr>
<tr>
<td>$q+1$</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob $q+1%$</td>
</tr>
<tr>
<td>$q+2$</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob $q+2%$</td>
</tr>
<tr>
<td>$q+3$</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob $q+3%$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>99</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ w/ prob 99%</td>
</tr>
<tr>
<td>100</td>
<td>$(3,0)$</td>
<td>or</td>
<td>$(20,20)$ 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
Issue 1: Consequentialism

- Elicit $u_i(15, 5)$, e.g.
- $u_i$ captures non-selfish preferences.
- $u_i$ captures risk aversion.

**Problem:** Game theory: utility over strategies: $U_i(s_i, s_j)$

We elicit: utility over outcomes: $u_i(\pi(s_i, s_j))$

**Solution:** Assume consequentialism:

$$U_i(s_i, s_j) = u_i(\pi(s_i, s_j))$$

Is consequentialism reasonable?? Is it even testable??
Example violating consequentialism:

<table>
<thead>
<tr>
<th></th>
<th>Nice</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Foolish</td>
<td>$5, $5</td>
<td>$5, $5</td>
</tr>
<tr>
<td>Wise</td>
<td>$100, $5</td>
<td>$5, $5</td>
</tr>
</tbody>
</table>

$\pi(\text{Foolish, Nice}) = \pi(\text{Wise, Mean})$, but, intuitively $U_1(\text{Foolish, Nice}) \neq U_1(\text{Wise, Mean})$.

But how could you possibly observe that??

**Claim:** Cannot elicit $U_i(s_i, s_j)$. *Must* assume consequentialism.

**Messy Solution:** Elicit $u_i(\pi(s_i, s_j))$ in the context of the game.
So, What Should We Measure?

2. Strategies: $s_i$

- Easy. Just play the game.
- Complete contingent plan
  - “When will you Take?”
- Can re-elicit this at each node
  - Even when not active
3. Beliefs: \( p_i^1(u_{-i}, s_{-i}) \), \( p_i^2(p_{-i}^1, u_{-i}, s_{-i}) \), \ldots

Measure before the game:

1. Best guess of \( u_{-i}(x_1, x_2) \) at each terminal node

Measure at every node:

1. Probability of each \( s_{-i} \) (call that \( p_i^1(s_{-i}) \))
2. Best guess of \( p_{-i}^1(s_i) \)
3. Probability \( -i \) is rational

My Wish List:

1. Entire distribution over \( u_{-i} \)
2. Correlation between \( u_{-i} \) and \( s_{-i} \)
3. Correlation between \( p_{-i}^1 \) and \( s_{-i} \)
Does elicitation contaminate game play? PROBABLY!
Does game play contaminate elicitation?? PROBABLY!

• I embrace it! This is a *fully contaminated* experiment!
  • Necessary evil for the methodology
  • Intuitively: provides an upper bound on rationality

Empirically, I think it probably doesn’t matter:

• In five $2 \times 2$ games, play w/out elicitation was the same in 4 of 5
• Behavior pretty similar to previous papers
You're about to choose PASS.
(you plan to choose DOWN at step #5).
Play will continue, with Player 2 choosing next.

CONFIRM & SUBMIT
For each step remaining for Player 2, indicate how likely you think it is they will choose PASS or DOWN, if that step is reached.

I confirm the above percentages are all as I want them: ☐
Shown below (in red) are your guesses of Player 2’s preferences and likelihood of you choosing PASS or DOWN at each step.

START
Player 1 PASS 80% Player 2

<table>
<thead>
<tr>
<th>Player 1</th>
<th>Player 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>PASS 45%</td>
<td>PASS 60%</td>
</tr>
<tr>
<td>DOWN 20%</td>
<td>DOWN 40%</td>
</tr>
</tbody>
</table>

P1: $12.00  P1: $17.00  P1: $16.00  P1: $22.00  P1: $21.00
P2: $9.00   P2: $15.00  P2: $14.00  P2: $20.00  P2: $25.00

is the same as 15% of $30 ea.

Your payoff is always shown in bold.

How likely do you think it is that Weighted Value Theory will correctly predict Player 2’s choices at each remaining step?

(Your guesses of their values and beliefs appear above.)
SUBJECT 316 (Player 2)

```
<table>
<thead>
<tr>
<th>Outcome</th>
<th>$24</th>
<th>$21</th>
<th>$22</th>
<th>$24</th>
<th>$25</th>
<th>$22</th>
<th>$23</th>
<th>$25</th>
<th>$26</th>
<th>$23</th>
<th>$24</th>
<th>$26</th>
<th>$27</th>
<th>$24</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>
```

```
<table>
<thead>
<tr>
<th>Choice</th>
<th>20%</th>
<th>60%</th>
<th>10%</th>
<th>45%</th>
<th>5%</th>
<th>40%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>80%</td>
<td>40%</td>
<td>90%</td>
<td>55%</td>
<td>95%</td>
<td>60%</td>
</tr>
<tr>
<td>2</td>
<td>80</td>
<td>70</td>
<td>80</td>
<td>75</td>
<td>85</td>
<td>80</td>
</tr>
<tr>
<td>1</td>
<td>70</td>
<td>80</td>
<td>70</td>
<td>80</td>
<td>80</td>
<td>75</td>
</tr>
</tbody>
</table>
```

P1Rat: 45%
Example Observation
Centipede Treatments

**CENT-LO:**
“Risk $1 to gain $5”

```
1 1 12 2 2
$12 $11 $17 $16 $22 $21
$ 9 $15 $14 $20 $19 $25
```

**CENT-HI:**
“Risk $2 to gain $1”

```
1 1 12 2 2
$24 $22 $25 $23 $26 $24
$21 $24 $22 $25 $23 $26
```

**CENT-ALL:**
“Risk ALL to gain $4”

```
1 1 12 2 2
$15 $ 1 $19 $ 1 $23 $ 1
$ 1 $15 $ 1 $19 $ 1 $23
```

$27 $24
$27 $24
$27 $ 1
Design Details

• OSU subject pool
• Custom software, ORSEE recruiting
• Between-subjects treatment (LO vs HI vs ALL)
• Play 4 periods. Elicitation only in last 2
  • Random rematching with feedback
• Only one game or elicitation is paid
• $19 average
• # subjects:

<table>
<thead>
<tr>
<th></th>
<th>CENT-LO</th>
<th>CENT-HI</th>
<th>CENT-ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>54</td>
<td>36</td>
<td>62</td>
</tr>
</tbody>
</table>
Results
CENT-LO: “Risk $1 to gain $5”
The Story

CENT-LO:
“Risk $1 to gain $5”

Outcome Frequencies (Last Period)

<table>
<thead>
<tr>
<th>Outcome Node</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
</tr>
</tbody>
</table>

Bar graph showing frequencies of outcomes from nodes 1 to 7 with node 7 having the highest frequency.
The Story

CENT-LO: “Risk $1 to gain $5”

1. Node 7 freq
2. Story???
CENT-LO: “Risk $1 to gain $5”

What do we learn from elicitation?

1. There are altruists who prefer Pass even if opponent will Take
   • Many people will give up $1 to give $6
2. Selfish people know that altruists are common
3. *Early nodes*: Selfish people Pass, knowing altruists Pass back
The Story

CENT-LO:
“Risk $1 to gain $5”
The Unit of Analysis: 3-Node Segments

A 3-Node Segment

<table>
<thead>
<tr>
<th>Node</th>
<th>1</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>$17</td>
<td>$16</td>
<td>$22</td>
</tr>
</tbody>
</table>

Selfish $u_1$

- $u_1 = 17, 16, 22$

<table>
<thead>
<tr>
<th>Node</th>
<th>1</th>
<th>2</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>$17</td>
<td>$16</td>
<td>$22</td>
</tr>
</tbody>
</table>

Altruist $u_1$

- $u_1 = 16, 17, 21$

SizeBAP: A measure of the temptation to Pass

- Let $p =$ subjective prob. next mover will Pass
- **Selfish:** Pass is BR if $p \in [1/6, 1]$
  - SizeBAP for this $u_1$ is 5/6. Very likely to Pass.
- **Altruist:** Pass is BR if $p \in [0, 1]$ (strict Dom.Strat.)
  - SizeBAP for this $u_1$ is 1. Guaranteed to Pass.
- SizeBAP is a statistic for $u_i$ (and nothing else)
Pooling All 3-Node Segments

CENT-LO Treatment:

Dominant Strategy (3-Node Game Segments)

SizeBAP Histogram

Game Outcome Frequencies (Final Period)

Selfish SizeBAP \approx 0.833
CENT-HI Treatment:

Dominant Strategy (3–Node Game Segments)

SizeBAP Histogram

Game Outcome Frequencies (Final Period)

Selfish SizeBAP $\approx 0.333$
CENT-ALL Treatment:

Dominant Strategy (3-Node Game Segments)

SizeBAP Histogram

Game Outcome Frequencies (Final Period)

Selfish SizeBAP ≈ 0.22, 0.18, 0.15
Verifying the Story: CENT-LO

CENT-LO:

1. Altruists exist
   • Pass is DomStrat in 43.7% of segments

2. Altruists pass
   • 89.7% of the time
   • $43.7\% \times 89.7\% = 39.2\%$ overall chance of Pass from altruists

3. Non-altruists believe Pass is reasonably likely
   • $54.8\%$ have $Pr(\text{Pass}) > 39.2\%$ (median = 40%)
   • Self-similarity hides direct belief in altruism

4. Non-altruists BR to that belief
   • $83.8\%$ play BR, given $p_i^1$ and $u_i$
Verifying the Story: CENT-HI

CENT-HI:

1. Altruists don’t exist
   • Pass is DomStrat in 8.9% of segments

2. Altruists pass but they’re very rare
   • Small sample: 6 out of 9
   • $8.9\% \times 66.6\% = 5.9\%$ overall chance of Pass

3. Non-altruists believe Pass is reasonably unlikely
   • Median = 20%
   • Self-similarity hides direct belief in altruism

4. Non-altruists BR to that belief
   • 58.5% play BR, given $p_i$ and $u_i$
   • Beliefs only elicited for those that Pass, which is a small sample
Verifying the Story: CENT-ALL

CENT-ALL:

1. Altruists don’t exist
   - Pass is DomStrat in 8.7% of segments

2. Altruists pass but they’re very rare
   - Small sample: 2 out of 12
   - $8.7\% \times 16.67\% = 1.45\%$ overall chance of Pass

3. Non-altruists believe Pass is reasonably unlikely
   - Median = 17.5%
   - Self-similarity hides direct belief in altruism

4. Non-altruists BR to that belief
   - 38.3% play BR, given $p_i$ and $u_i$
   - Beliefs only elicited for those that Pass, which is a small sample
Belief in Rationality & Backward Induction

• Does common belief in rationality ⇒ backwards induction?
• Depends how people react to surprises (Reny 1993)
  • RCSBR: continue to believe in rationality after surprises
  • (Surprises ⇒ belief in irrationality) ⇒ Surprises!
• Surprise: Pr(Take)=100%, Pr(Rational)=100%, but then Pass
Belief in Rationality & Backward Induction

- Does common belief in rationality ⇒ backwards induction?
- Depends how people react to surprises (Reny 1993)
  - RCSBR: continue to believe in rationality after surprises
  - (Surprises ⇒ belief in irrationality) ⇒ Surprises!
- Surprise: Pr(Take)=100%, Pr(Rational)=100%, but then Pass
- **CENT-LO**: Pr(Take) never near 100%
  - It’s *not* a game of complete information!
Belief in Rationality & Backward Induction

- Does common belief in rationality $\Rightarrow$ backwards induction?
- Depends how people react to surprises (Reny 1993)
  - RCSBR: continue to believe in rationality after surprises
  - (Surprises $\Rightarrow$ belief in irrationality) $\Rightarrow$ Surprises!
- Surprise: $\Pr(\text{Take})=100\%, \ Pr(\text{Rational})=100\%$, but then Pass
- CENT-ALL: Very few surprises since everyone Takes!
  - Unsurprisingly, surprises are rare
Prisoners Dilemma
The Prisoners’ Dilemma

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$10, $10</td>
<td>$1, $15</td>
</tr>
<tr>
<td>D</td>
<td>$15, $1</td>
<td>$5, $5</td>
</tr>
</tbody>
</table>

The Prisoners’ Dilemma Game Form

- New treatment: SIM
- Five $2 \times 2$ games without feedback, random matching
- Elicitation in every game
- Pencil & paper
- $n = 150$
# The Prisoners' Dilemma

The Prisoners' Dilemma Game Form

<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>C</td>
<td>$10, $10</td>
<td>$1, $15</td>
</tr>
<tr>
<td>D</td>
<td>$15, $1</td>
<td>$5, $5</td>
</tr>
</tbody>
</table>

30.4% play C.

Why???
### The Prisoners' Dilemma

The Prisoners' Dilemma Game Form

| Pref. Type       | $BR_i(C)$ | $BR_i(D)$ | % Subj. | $BR_i(p_i^{1s}|u_i) = C$ | $BR_i(p_i^{1s}|u_i) = D$ |
|------------------|-----------|-----------|---------|--------------------------|--------------------------|
|                  |           |           |         | $s_i = C$                | $s_i = D$                |
|                  |           |           |         | $s_i = C$                | $s_i = D$                |
| Selfish          | $D$       | $D$       | 68.0%   | –                        | 18                       |
| Cond. Coop.      | $C$       | $D$       | 19.7%   | 15                       | 5                        |
| Reverse          | $D$       | $C$       | 2.7%    | 1                        | 2                        |
| Uncond. Coop.    | $C$       | $C$       | 9.5%    | 8                        | 6                        |

Only 53% of cooperation (C) is rational. Failure of consequentialism or dominance.
### The Prisoners’ Dilemma

The Prisoners’ Dilemma Game Form

| Pref. Type        | $BR_i(C)$ | $BR_i(D)$ | % Subj. | $BR_i(p_i^{1s}|u_i) = C$ | $BR_i(p_i^{1s}|u_i) = D$ |
|-------------------|-----------|-----------|---------|--------------------------|--------------------------|
| Selfish           | $D$       | $D$       | 68.0%   | –                        | 18                       |
| Cond. Coop.       | $C$       | $D$       | 19.7%   | 15                       | 3                        |
| Reverse           | $D$       | $C$       | 2.7%    | 1                        | 0                        |
| Uncond. Coop.     | $C$       | $C$       | 9.5%    | 8                        | –                        |

Only 53% of cooperation (C) is rational

Failure of consequentialism or dominance
Iterated Dominance
A Dominance-Solvable Game Form

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>$10, $5</td>
<td>$15, $15</td>
</tr>
<tr>
<td>D</td>
<td>$5, $10</td>
<td>$1, $1</td>
</tr>
</tbody>
</table>

A Dominance Solvable Game Form

- **Row players:** 100% play U
  - 71 of 75: U is a dominant strategy
  - 4 of 75: U is a best response
- **Column players:** 25% play L
  - Why???
A Dominance-Solvable Game Form

| Pref. Type | $BR_i(U)$ | $BR_i(D)$ | % Subj. | $BR_i(p_i^{1s}|u_i) = L$ | $BR_i(p_i^{1s}|u_i) = R$ |
|------------|-----------|-----------|---------|--------------------------|--------------------------|
| Selfish    | $R$       | $L$       | 91.9%   | 0                        | 14                       |
| DomStrat L | $L$       | $L$       | 5.4%    | 3                        | 1                        |
| DomStrat R | $R$       | $R$       | 2.7%    | –                        | 1                        |
| Reversed   | $L$       | $R$       | 0%      | 0                        | 0                        |

Violation of consequentialism and/or EU
Conjecture: avoiding ($1, $1), despite stated preferences. Strategic uncertainty.
SEQ treatment: $n = 60$

Irrationality *disappears* when strategic uncertainty is removed.
Coordination
An Asymmetric Coordination Game Form

- Row: 93% play U
- Col: 49.3% play L
  - Why??? Beliefs?
# Asymmetric Coordination

## An Asymmetric Coordination Game Form

| Row’s Type | $BR_1(L)$ | $BR_1(R)$ | % Subj. | $BR_1(p_1^{US}|u_1) = U$ | $BR_1(p_1^{US}|u_1) = D$ |
|------------|-----------|-----------|---------|--------------------------|--------------------------|
| Selfish    | $U$       | $D$       | 95.8%   | 43                       | 20                       |
| DomStrat U | $U$       | $U$       | 4.2%    | 3                        | –                        |

| Col’s Type | $BR_2(U)$ | $BR_2(D)$ | % Subj. | $BR_2(p_2^{US}|u_2) = L$ | $BR_2(p_2^{US}|u_2) = R$ |
|------------|-----------|-----------|---------|--------------------------|--------------------------|
| Selfish    | $L$       | $R$       | 93.0%   | 26                       | 27                       |
| DomStrat L | $L$       | $L$       | 7.0%    | 5                        | 0                        |

Payment Matrix:

<table>
<thead>
<tr>
<th></th>
<th>L</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>$15, 5$</td>
<td>$2, 1$</td>
</tr>
<tr>
<td>D</td>
<td>$1, 2$</td>
<td>$5, 10$</td>
</tr>
</tbody>
</table>
Lessons

1. Most experiments are Bayesian games, not complete info
2. The story changes from one game to the next
3. Centipede game forms:
   • Altruists pass $\Rightarrow$ selfish pass
   • Backwards induction seems to work fine here
4. Prisoners’ dilemma:
   • Non-consequential preference for cooperating
5. Beliefs are generally pretty accurate
6. Don’t write a solo-authored paper post-tenure