Incentives in Experiments: Theory and an Experimental Test

Theory:
Yaron Azrieli (OSU), Chris Chambers (UCSD), P.J. Healy (OSU)
Experiment:
Alex Brown (Texas A&M), P.J. Healy (OSU)

March 2015
UC Santa Barbara

How Do Experimenters Pay Subjects?

2011 Publications:

Payment:	Only 1 Task	None Paid	One Random	Some Random	All Paid	Rank-Based	Total
	Individual Choice Experiments						
' Top 5 '	7	0	3	1	3	0	14
Exp.Econ.	3	0	1	0	2	0	6
	Muti-Person (Game) Experiments						
' Top 5 '	9	0	1	0	8	0	18
Exp.Econ.	8	1	3	3	5	1	21
Total	27	1	8	4	18	1	59

LESSON: There is no convention on how to pay subjects.

Problematic Experiments

Pay-All Mechanism

1. Problem 1: {beer,milk}
2. Problem 2: {hot dog,chocolate cake}

1. Any normal human: beer$\succ$ milk, and cake$\succ$ hot dog
2. TRUTH: (beer,cake) $\rightarrow$ {beer,chocolate cake}
3. LIE: (beer,hot dog) $\rightarrow$ {beer,hot dog}
4. Any normal human: LIE$\succ$ TRUTH (Not "incentive compatible")

Other ways it can fail:

1. Wealth effects
2. Portfolio effects
3. Hedging incentives
4. *ex post* fairness concerns

Problematic Experiments

Random Problem Selection (RPS) Mechanism ('pay one randomly')
Let $L = (\frac{1}{2}, \$0; \frac{1}{2}, \$3)$.

1. Problem 1: $\{L, \$1\}$
2. Problem 2: $\{L, \$2\}$

1. Subject: $L \succ \$1$, and $\$2 \succ L$.
2. TRUTH: $(L, \$2) \rightarrow (\frac{1}{2}, L; \frac{1}{2}, \$2) \longrightarrow^{\text{Red.}} (0.25, \$0; 0.5, \$2; 0.25, \$3)$
3. LIE: $(\$1, \$2) \rightarrow (0.5, \$1; 0.5, \$2)$
4. $\exists$ rank-dependent utility prefs. where LIE $\succ$ TRUTH

Other ways it can fail:

1. Ambiguity aversion
2. *ex ante* fairness concerns

Choice Objects vs. Payment Objects

Pay-All Mechanism:

1. Problem 1: $\{$beer,milk$\}$, Problem 2: $\{$hot dog,chocolate cake$\}$
2. Choice objects: $X = \{$beer,milk,hot dog,chocolate cake$\}$
3. Payment objects: $P(X)$=$\{\{$beer,hot dog$\},\{$beer,cake$\},\{$milk,hot dog$\},\{$milk,cake$\}\}$

RPS Mechanism:

1. Problem 1: $\{L, \$1\}$, Problem 2: $\{L, \$2\}$
2. Choice objects: $X = \{$simple lotteries$\}$
3. Payment objects: $P(X) = \{$compound lotteries$\}$

LESSON: Incentives depend on $\succeq$ over $P(X)$, not X

Are Experimenters Making A Mistake?

- Experimenters interested in $\succeq$ over X (choices).
- Suppose they have theory/hypotheses about $\succeq$ on X.
- If theory does not extend to $P(X)$, then we cannot judge incentive properties of experiment!

How many experimenters are being careful about $P(X)$ vs. X?

Discussion of Incentives

The 31 papers from 2011 with multiple problems given:

	Mechanism Not in Paper	Discussion of Incentives None	Brief	Extensive	Clearly I.C.	Total
		Individual Choice Experiments				
'Top 5'	1	6	0	1	0	7
Exp.Econ.	0	2	0	1	0	3
		Muti-Person (Game) Experiments				
'Top 5'	6	9	0	0	0	9
Exp.Econ.	2	7	4	1	0	12
Total	9	24	4	3	0	31

What We Are Doing

Goal of theory paper w/ Azrieli & Chambers:
Understand what assumptions about $P(X)$ make each mech. I.C.

Goal of experimental paper w/ Brown:
Test I.C. of a popular experimental protocol

The Literature

- Invention of pay-one-randomly ('RPS') & I.C. under SEU:
  Wold (1952), Savage (1954), Allais (1953), Wallis.
- RPS not I.C. with non-EU:
  Holt (1986), Karni & Saffra (1987)
- Experiments showing RPS works:
  Camerer (1989), Loomes et al. (1991), Starmer & Sugden (1991), Beatte & Loomes (1997), Cubitt et al. (1998)
- Justification via prospect theory:
  Wakker et al. (1994), Cubitt et al. (1998)
- Experiments showing RPS fails:
  Cox et al. (2014a,b), Harrison & Swarthout (2014)
- Not IC with ambiguity:
  Baillon et al. (WP), Oechssler & Roomets (2014)

What is an Experiment?

Take the viewpoint of a single subject.

An experiment consists of:

1. List of decisions to be made
2. A payment rule

Researcher's objective: observe choice function over given decisions

(**Our objective:** avoid the 'theory-reality' gap)

The Setup

Formally:

- Decision problems: $D = (D_1, \ldots, D_k)$
  - $D_i \subseteq X =$ 'choice objects'. No structure.
  - X, k finite
- Choice: $\succ$ over X (complete & transitive). This talk: strict.
  - $\mu_i(\succ) = \{x \in D_i : (\forall y \in D_i)\ x \succ y\}$
    "*True* favorite from D_i"
- Payment Mechanism: ϕ
  - Messages: $M = \times_i D_i$ ('announced choice')
  - Mechanism: $\phi : M \to P(X)$ ($P(X)$ is TBD)
  - Payment: $\phi(m) \in P(X)$

Experiment: (D, ϕ)

An Example

Hypothesis: Dictator game giving correlates with risk preferences.

- First: 5-question Holt-Laury elicitation
  $D_1 = \{(0.1, \$2; \$1.60), (0.1, \$3.85; \$0.10)\}$. $m_1 = (0.1, \$2; \$1.60)$
  $D_2 = \{(0.3, \$2; \$1.60), (0.3, \$3.85; \$0.10)\}$. $m_2 = (0.3, \$2; \$1.60)$
  $D_3 = \{(0.5, \$2; \$1.60), (0.5, \$3.85; \$0.10)\}$. $m_3 = (0.5, \$2; \$1.60)$
  $D_4 = \{(0.7, \$2; \$1.60), (0.7, \$3.85; \$0.10)\}$. $m_4 = (0.7, \$3.85; \$0.10)$
  $D_5 = \{(0.9, \$2; \$1.60), (0.9, \$3.85; \$0.10)\}$. $m_5 = (0.9, \$3.85; \$0.10)$
- Next decision: dictator game
  $D_6 = \{(\$100, \$0), (\$99, \$1), \ldots, (\$0, \$100)\}$. $m_6 = (\$90, \$10)$
- RPS mechanism w/ 6-sided die: Roll j, pay m_j
- Pay-all mechanism: Pay $\{m_1, m_2, \ldots, m_6\}$
- Mixed mechanism w/ 5-sided die: Roll j, pay $\{m_j, m_6\}$.

Payment Objects

What are the possible payment objects?

- Bundles: $B(X) = \{\{m_1, \ldots, m_k\} : m_i \in D_i \ \forall i\}$
- Randomizing Device: $\Omega = \{\omega_1, \ldots, \omega_t\}$
  Conditional payment (Savage act): $\Omega \mapsto B(X)$

Payment objects: $P(X) = B(X)^{\Omega}$

- Pay-all mechanism: $\Omega = \{\omega\}$, $\phi(m)(\omega) = \{m_1, \ldots, m_k\}$
  So $P(X) = B(X)$
- RPS mechanism: $\Omega = \{\omega_1, \ldots, \omega_k\}$, $\phi(m)(\omega_i) = m_i$
  So $P(X) = X^{\Omega}$.

Preference $\succ$ over X extends to $\succ^*$ over $P(X)$

I.C.: (D, ϕ) is IC if $\phi(\mu(\succ)) \succ^* \phi(m)$ $\forall m \neq \mu(\succ)$.

What should we assume about $\succ^*$?

Consistency: If $\phi(m)$ pays x in every state, and $\phi(m')$ pays y in every state, then $\phi(m) \succ^* \phi(m') \Leftrightarrow x \succ y$.

Theorem

If only consistency is assumed, then IC $\Leftrightarrow$ only one task.

"IC is never free."

Paying in Bundles

Pay-all: $\Omega = \{\omega\}$.
$\phi(m)(\omega) = \{m_1, m_2, \ldots, m_k\} \leftarrow$ 'bundle' payment

No Complementarities at the Top (NCaT): For every $m \neq \mu(\succ)$,

$$\{\mu_1(\succ), \ldots, \mu_k(\succ)) \succ^* \{m_1, m_2, \ldots, m_k\}$$

"Bundle of favorites is your favorite bundle."

Theorem

Assume NCaT and nothing else. Then ϕ is IC if and only if it is 'equivalent to' the pay-all mechanism.

In almost all applications, 'equivalent to' means 'equals'.

Complementarities

Examples of complementarities:

- Suppose each $D_i = \{x_i, y_i\}$, where x_i is safe, y_i is risky.
  - Wealth effect: $x_1 \succ y_1$ but $\{\$1000000, y_1\} \succ^* \{\$1000000, x_1\}$.
  - Portfolio effect: $x_i \succ y_i \ \forall i$, but $\{y_1, \ldots, y_k\} \succ^* \{x_1, \ldots, x_k\}$.
- Consumption goods: beer-hot dog example
- Fairness: \$10 Gary$\succ$ \$10 Ryan, but $\{$\$10 Gary,\$10 Ryan$\} \succ^* \{$\$20 Gary$\}$.

Monotonicity

Consider RPS mechanism and $(D_1, \ldots, D_k)$.
Suppose $\mu_i(\succ) = x_i$ in every D_i.

	Die Roll ('state')					
Strategy	1	2	3	4	$\cdots$	k
Truth:	x_1	x_2	x_3	x_4	$\cdots$	x_k
Lie:	x_1	y_2	y_3	x_4	$\cdots$	x_k

Monotonicity: If $\phi(m)(\omega) \succ \phi(m')(\omega)$ $\forall \omega$ then $\phi(m) \succ^* \phi(m')$.

Theorem

If $\succ^*$ satisfies monotonicity, then RPS is IC.

Examples where RPS fails must violate monotonicity.

Monotonicity & RPS

Theorem

If we assume monotonicity and nothing more,
then RPS is essentially the only IC mechanism.

Example of a non-RPS IC mechanism

- $D_1 = \{x, y\}$, $D_2 = \{y, z\}$, $D_3 = \{x, z\}$.
- RPS:

State:	ω_1	ω_2	ω_3
$\phi(m)(\omega)$:	m_1	m_2	m_3

- Rationalizable: $m = (x, y, x)$ reveals $x \succ y \succ z$.
- Non-rationalizable: $m' = (x, y, z)$ reveals $x \succ y \succ z \succ x$!
- If m is rationalizable:

State:	ω_1	ω_2	ω_3	ω_4
$\phi(m)(\omega)$:	m_1	m_2	m_3	revealed favorite in $\{x, y, z\}$

- If m not rationalizable:

State:	ω_1	ω_2	ω_3	ω_4
$\phi(m)(\omega)$:	x	y	x	z

Problem: Rarely can this be done in practice. So, RPS is it.

On Monotonicity

Monotonicity is weak on its own...
But becomes strong with other axioms!

On Monotonicity & Reduction

- Suppose $\{\omega_1, \ldots, \omega_k\} \mapsto (p_1, \ldots, p_k)$ (prob. sophistication)
- RPS mechanism creates lottery $(p_1, m_1; p_2, m_2; \ldots; p_k, m_k) \in P(X)$.
- Monotonicity: $x_1 \succ y_1 \Rightarrow (\alpha, x_1; 1-\alpha, z_2) \succ^* (\alpha, y_1; 1-\alpha, z_2)$
  Linear indifference curves in $P(X)$!
- The 'reduced mixture' is $\sum_i p_i \cdot m_i \in X$. (Need X convex.)

Reduction:

$(p_1, m_1; \ldots; p_k, m_k) \succ^* (p'_1, m'_1; \ldots; p'_k, m'_k) \iff \sum_i p_i m_i \succ \sum_i p'_i m'_i$

Basically says $\succ = \succ^*$.

Observation

Monotonicity + Reduction $\Rightarrow$ *linear indifference curves on* X.

On Monotonicity & Reduction

If X is space of simple lotteries:

Observation

Monotonicity + Reduction $\Rightarrow \succ$ on X is expected utility.

If X is space of (possibly ambiguous) acts:

Observation

Monotonicity + Reduction $\Rightarrow \succ$ on X is ambiguity neutral.

LESSON: RPS probably not IC under reduction.

The Bipolar Behaviorist

What about Holt, Karni-Safra, Cox *et al.*, Harrison & Swarthout...?

The Bipolar Behaviorist Claim

You cannot test non-EU theories using the RPS mechanism.

Our Claim

If you want to test non-EU theories using the RPS mechanism, you need to assume reduction is violated.

Evidence for reduction: mostly against.

Summary of Theory

- No assumptions: IC $\iff$ one task.
- NCaT: IC $\iff$ pay-all.
- Monotonicity: IC $\iff$ RPS
  - Monotonicity + Reduction $\Rightarrow$ linear indiff. on X.
- Pay multiple randomly: mix of monotonicity & NCaT.

How To Test for IC

Given *any* experiment (D, ϕ), we can test for IC directly:

- Recruit large number of subjects.
- Randomly split into $(k+1)$ treatments.
- Treatment 0: Original experiment $((D_1, \ldots, D_k), \phi)$.
- Treatment i: "ith IC-testing treatment" $((D_1, \ldots, D_k), \phi_i^I)$, where
  $\phi_i^I(m) = m_i$ for some fixed i.
- $\rho_i^i(x)$ is frequency of subjects choosing $x \in D_i$ in Trt i.
- $\rho_i^0(x)$ = frequency of subjects choosing $x \in D_i$ in Trt 0.
- k Fisher (or χ^2) tests for $\rho_i^i \equiv \rho_i^0$.
  - Assumes independence across D_i. Or, do Bonferoni-type correction.
- Controls framing effects.
- Power is not good.
- Can test "IC on D_i" by doing Trt 0 vs Trt i only.

Evidence

Shift to experimental designs & data.

Heavily Confounded Tests

- Cox Sadiraj & Schmidt (2014a)
  - Trt 1: $\{\$4, (\frac{1}{2}, \$10)\}$
  - Trt 2: $\{\$3, (\frac{1}{2}, \$12)\}$ and $\{\$4, (\frac{1}{2}, \$10)\}$. RPS.
  - Decoy effect causes framing effect. NOT a violation of IC.
- Cubitt Starmer Sugden (1998 Exp.1)
  - Trt 1: $(D_1, \ldots, D_{18}, D_{19}, D_{20})$. RPS on 19,20. $n = 57$.
  - TRT 2: $(D_1, \ldots, D_{18}, D_{19}, D'_{20})$. RPS on 19,20'. $n = 62$.
  - Test on D_{19}: 0.924. No framing.

Framing-Confounded Tests

- Beattie & Loomes (1997)
  - Group 0: (D_1, D_2, D_3, D_4). Paid via RPS. $n = 49$.
  - Group i: (D_i). Paid for D_i. $n = 48$.
  - p-values: 0.36, 0.82, 0.74, 0.064. (Do reject reduction.)
- Cubitt Starmer Sugden (1998 Exp.2)
  - Group 0: $(D_1, \ldots, D_{18}, P', P'')$. RPS over P', P''. $n = 51$.
  - Group 1: $(D_1, \ldots, D_{18}, D_{19}, P')$. Paid only P'. $n = 53$.
  - p-value: 0.720.

Framing-Confounded Tests

- Cox Sadiraj & Schmidt (2014b)
  - Trt 0: $(D_1, \ldots, D_5)$. RPS. $n = 40$
  - Trt i: (D_i). Pay only D_i. $n = 46.2$ avg.
  - p-values: 0.24, $<$0.001, 0.15, 0.50, 0.28
  - Similar results for 3 variations on RPS.
- Harrison & Swarthout (2014)
  - Trt 0: $(D_1, \ldots, D_{30})$. RPS. $n = 208$.
  - Trt R: One randomly-chosen D_i. Paid only D_i. $n = 75$.
  - Estimate RDU functions, adding demographics.
  - Estimates differ between Trt 0 and Trt R.

Starmer Sugden (1991):

- $P' = \{S', R'\}$. $P'' = \{S'', R''\}$.
- Group A: $(D_1, \ldots, D_{20}, P', P'')$. Paid only for P''. $n = 40$.
- Group B: $(D_1, \ldots, D_{20}, P', P'')$. RPS over P' and P'' only. $n = 40$.
- Group C: $(D_1, \ldots, D_{20}, P'', P')$. RPS over P'' and P' only. $n = 40$.
- Group D: $(D_1, \ldots, D_{20}, P'', P')$. Paid only for P''. $n = 40$.

Their test p-values:

- P'' in A vs (B+C): 0.223
- P' in D vs (B+C): 0.0516
- Results borderline. Admit low power. (Reduction violated.)

Our tests (no framing confound):

- P'' in A vs B: 0.356
- P' in D vs C: 0.043

Cubitt Starmer Sugden (1998, Experiment 3

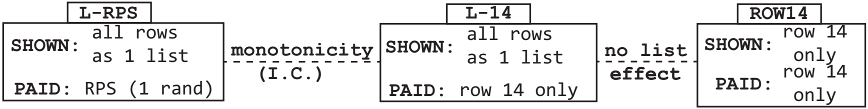
- Group A: $(D_1, D_2, D_3, \ldots, D_{20})$. Paid only D_1. $n = 49$.
- Group B: $(D_1, D_2, D_3, \ldots, D_{20})$. Paid only D_2. $n = 56$.
- Group C: $(D_1, D_2, D_3, \ldots, D_{20})$. RPS over 1,2. $n = 52$
- p-values: A vs C: 0.685. B vs C: 0.120.

Why We Need Another Test

Issues with existing tests.

1. Confound with framing (except SS91 and CSS98 Exp3)
2. Lack of power
3. Oversampling common-ratio test lotteries, etc.
4. Added hypothetical questions (piloting).

Our Design



- Holt-Laury questions
- Andreoni-Sprenger formatting
- Standard Ohio State subject pool.
- Between-subjects.
- Computerized.
- Physical randomizing devices (die, bingo cage)
- No other tasks in the experiment.
- 60–63 subjects per treatment.
- List format: rows must be answered sequentially.

The List

Row #	Option A		or	Option B	
1	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	or	Ball 1 pays $15 (5% chance of $15)	Balls 2-20 pay $0 (95% chance of $0)
2	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	or	Balls 1-2 pay $15 (10% chance of $15)	Balls 3-20 pay $0 (90% chance of $0)
3	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	or	Balls 1-3 pay $15 (15% chance of $15)	Balls 4-20 pay $0 (85% chance of $0)
4	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	or	Balls 1-4 pay $15 (20% chance of $15)	Balls 5-20 pay $0 (80% chance of $0)
	Balls 1-10 pay $10	Balls 11-20 pay $5		Balls 1-5 pay $15	Balls 6-20 pay $0
⋮	⋮			⋮	⋮
18	(50% chance of $10)	(50% chance of $5)	or	(90% chance of $15)	(10% chance of $0)
19	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	or	Balls 1-19 pay $15 (95% chance of $15)	Ball 20 pays $0 (5% chance of $0)
20	Balls 1-10 pay $10 (50% chance of $10)	Balls 11-20 pay $5 (50% chance of $5)	or	All Balls pay $15 (100% chance of $15)	(0% chance of $0)

Click Here When Finished

The Results

- Showing whole list makes them switcher earlier (Closer to the middle.)
  - Not quite significant.
- Using RPS mechanism makes them switch later. (More thoughtful? Switching inertia?)
  - Statistically significant.

Hypothesis

Hypothesis

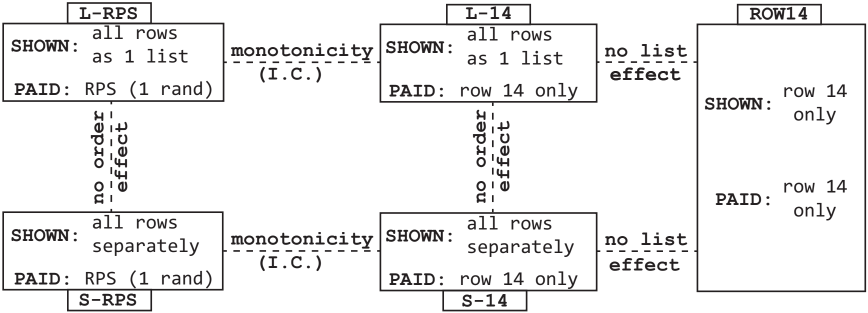
- *Subjects are combining the decisions in a reduction-like way. E.g.: 'When to switch?'.*
- *The 'combining' can be broken by separating the decisions.*

New Treatments

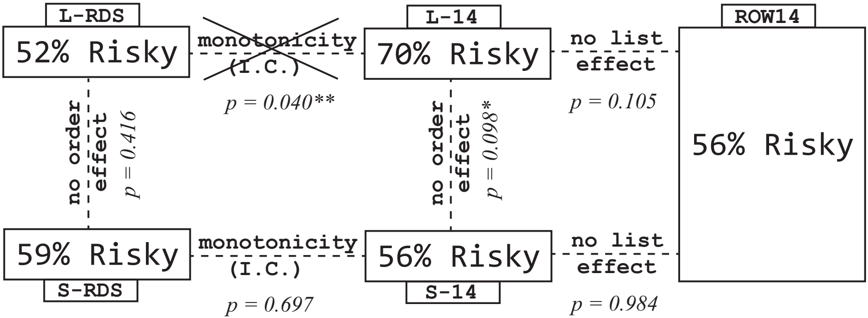
'Separated' treatments.

- Same 20 rows.
- Each shown on separate screen.
- Order randomized for each subject.
- Still comparing RPS to Pay-14-Only.
- Still must answer every row, in order given.
- Still 60–63 observations per cell, between subjects.

Full Design


L-RPS
SHOWN: all rows as 1 list
PAID: RPS (1 rand)
monotonicity (I.C.)
L-14
SHOWN: all rows as 1 list
PAID: row 14 only
no list effect
ROW14
SHOWN: row 14 only
PAID: row 14 only
no order effect
no order effect
SHOWN: all rows separately
PAID: RPS (1 rand)
S-RPS
monotonicity (I.C.)
SHOWN: all rows separately
PAID: row 14 only
S-14
no list effect

The Results


L-RDS
52% Risky
monotonicity (I.C.)
p = 0.040**
L-14
70% Risky
no list effect
p = 0.105
ROW14
56% Risky
no order effect
p = 0.416
no order effect
p = 0.098*
59% Risky
S-RDS
monotonicity (I.C.)
p = 0.697
56% Risky
S-14
no list effect
p = 0.984

The Cost of Separation

B-to-A switches are an indirect violation of monotonicity in $\succ$ (not $\succ^*$).

$Risky_{15}$ dominates $Risky_{14}$, but $Risky_{14} \succ Safe \succ Risky_{15}$

# B-to-A Switches	L-RPS (List)	S-RPS (Separated)
Zero	95.0%	67.2%
One	0%	29.5%
Two	0%	0%
Three	1.7%	3.3%
Four or more	3.3%	0%
χ^2 p-value	*0.00013****	

Additional Results

1. Hypothesis: B-A switches occur 'later'
   - Result: they occur earlier!
   - 4.5% in first choice, 1.5% in last choice.
2. List-Framing result does become significant in regressions controlling for gender & Big-5.

To-Do List:

- Speed of decisions.
- Other suggestions?

Summary

- RPS can fail.
- RPS has its best shot when decisions are separated.
- Separation may come at a cost.
- Future work: How to minimize inconsistency with separation?
  - Question: should we be ‘forcing’ consistency?

Thank You.