# ExpEcon Methods:
# Empirical Tests of Incentive Compatibility

ECON 8877

P.J. Healy

Updated 2026-01-28 at 01:34:56

# Testing IC vs. Framing Effects

A test of IC? (Cox Sadiraj & Schmidt 2014)

| | $D_1$ | $D_2$ |
|---|---|---|
| Treatment 1: | $\{\$4, (\frac{1}{2}, \$10)\}$ | |
| Treatment 2: | $\{\$4, (\frac{1}{2}, \$10)\}$ | $\{\$3, (\frac{1}{2}, \$12)\}$ |

If we observe differences on $D_1$, it could be

- the mechanism was not IC, or
- the presence of $D_2$ altered preferences (e.g., decoy effect).

Other papers that use this method:

- Cubitt Starmer Sugden (1998 Exp.1)
- Beattie & Loomes (1997)
- Cubitt Starmer Sugden (1998 Exp.2)
- Harrison & Swarthout (2014)
- Cox Sadiraj & Schmidt (2015)

# Tests Without Framing Confound

Replace Treatment 1 with a "Framed Control" treatment:

| | $D_1$ | $D_2$ | Mechanism |
|---|---|---|---|
| Treatment 1: | $\{\$4, (\frac{1}{2}, \$10)\}$ | $\{\$3, (\frac{1}{2}, \$12)\}$ | Pay only $D_1$ |
| Treatment 2: | $\{\$4, (\frac{1}{2}, \$10)\}$ | $\{\$3, (\frac{1}{2}, \$12)\}$ | RPS |

**LESSON: Proper test of IC must show all subjects same choices.**

# Cox, Sadiraj & Schmidt (2015)

Test various payment mechanisms in lottery choice setting

1. Pay All (PA)
    - PAS: Sequentially (learn outcome each period)
    - PAI: Independently at the end
2. Pay One Randomly (POR)
    - 2.1 PORpi: with prior info about all choices to be made
    - 2.2 PORnp: no info about upcoming choices
    - 2.3 PORpas: learn realized payoffs you go, then get 1 at the end
3. Pay All Correlated (PAC) (lotteries must have same state space)
    - PAC/N divides payoffs by # of decisions, to match POR
4. One Task (OT)
    - 4.1 ImpureOT: Make all choices, but only one is paid
        - Added by a referee (not me!) and reported separately

# Cox, Sadiraj & Schmidt (2015)

Design:

- Choice over 5 lottery pairs
- Testing various versions of Allais paradox
- OT: between-subjects. All others: within-subject
  - Therefore OT Allais paradoxes are between-subject via Probit
- Choices on 5 separate slips of paper in an envelope

Analyses:

- Probit on Pr(Allais paradox) including demographics, EV, etc.
- Choice frequencies
- Probit on choice frequencies

# Cox, Sadiraj & Schmidt (2015)

**Table 3** Test results for Hypotheses 1–4

| Mechanism | CRE | CCE | DCRE | DCCE |
|---|---|---|---|---|
| OT | No | No | No | Yes[e] |
| PORnp | No | No | No | No |
| PORpi | No | Yes[b] | Yes[c] | No |
| PORpas | Yes[a] | No | No | No |
| PAS | No | Yes[b] | No | No |
| PAI | Yes[a] | No | No | No |
| PAC/N | No | Yes[b] | No | No |
| PAC | Yes[a] | No | Yes[d] | No |

Notes: [a]Fan Out; [b]Fan In; [c]IRRA; [d]DRRA; [e]IARA

Can't really compare cleanly to OT
But, definite differences across mechanisms
And whether they see the questions in advance or not!

# Cox, Sadiraj & Schmidt (2015)

**Table 4** Observed frequencies (in %) of choices of less risky options (low and high column figures in bold)

| Mechanism | $S_1$ | $S_2$ | $S_3$ | $S_4$ | $S_5$ | All Pairs [95 % CI] |
|---|---|---|---|---|---|---|
| OT (231 subjects) | **39.47** | **15.52** | 27.59 | 28.95 | 38.46 | 28.60 [22.7, 34.4] |
| PORnp (40 subjects) | 37.50 | 45.00 | **47.50** | 32.50 | **60.00** | 44.50 [37.6, 51.4] |
| PORpi (40 subjects) | 27.50 | 50.00 | 42.50 | 22.50 | 50.00 | 38.50 [31.7, 45.3] |
| PORpas (40 subjects) | **22.50** | 42.50 | **20.00** | **10.00** | 30.00 | 25.00 [18.9, 31.1] |
| PAS (39 subjects) | 25.64 | 23.08 | 33.33 | 10.26 | **17.95** | 22.10 [16.2, 27.9] |
| PAC (38 subjects) | 36.84 | **52.63** | 23.68 | 21.05 | 42.11 | 35.30 [28.4, 42.1] |
| PAC/N (40 subjects) | 37.50 | 35.00 | 35.00 | 22.50 | 45.00 | 35.00 [28.3, 41.7] |
| PAI (38 subjects) | 36.84 | **52.63** | 36.84 | **34.21** | 52.63 | 42.60 [35.5, 49.7] |

All Pairs: % who chose safe in all 5 Most risk averse: PORnp and PAI
Least risk averse: PORpas and PAS

# Cox, Sadiraj & Schmidt (2015)

What about Impure OT?

- Paper only compares Impure OT to OT
    - More risky choices under Impure OT
    - Framing effect exists!
- But we want Impure OT vs. each mechanism!
- Probit Pr(Safe) results:
    - PORnp, PORpi, and PAI are different from ImpureOT
- But, looking at the actual choice data task-by-task, I don't find significant differences...

# Starmer & Sugden (1991)

- 22 binary lottery choice questions. $n = 40$ per treatment
- First 20: hypothetical (piloting for another study)
- Questions 21 and 22: RPS vs. only one paid. Same page.
- Allais paradox questions.

TABLE 1—THE DESIGN OF THE EXPERIMENT

| Group | Question 21 | Question 22 | Incentive |
|---|---|---|---|
| A | P′ | P″ | P″ is for real |
| B | P′ | P″ | Each problem has 0.5 chance of being for real |
| C | P″ | P′ | Each problem has 0.5 chance of being for real |
| D | P″ | P′ | P′ is for real |

A vs. B: $p = 0.356$ (my calculation)
C vs. D: $p = 0.043$ (my calculation)

# Cubitt Starmer & Sugden (1998)

- Five binary menus of lotteries
- Experiment 1 ($n = 201$)
    - Group 1.1: RPS: $(1/3, D_3; 2/3, D_4)$
    - Group 1.2: RPS: $(1/3, D_3; 2/3, D_5)$
    - (Two other groups to test IND and ROCL)
    - Use $D_3$ to test IC. No differences.
- Experiment 3 ($n = 202$)
    - 3.1: 20 decisions, 1st is paid
    - 3.2: 20 decisions, 2nd is paid
    - 3.3: 20 decisions, RPS on all 20
    - 3.4: Same as 3.3 but with lower stakes
    - 3.1 $D_1$ vs 3.3 $D_1$: $p = 0.685$
    - 3.2 $D_2$ vs 3.3 $D_2$: $p = 0.120$

# Summary of Past Experiments

**Table 9**
Existing tests of incentive compatibility of the RPS mechanism that have no framing confounds. We describe each of these comparisons in the text below.

| Paper | Names of treatments | [redacted] | $p$-Value | RPS is I.C.? |
|---|---|---|---|---|
| Starmer and Sugden (1991) | A vs. B | [redacted] | 0.356 | ✓ |
| Starmer and Sugden (1991) | C vs. D | [redacted] | **0.043** | × |
| [redacted] | | | | |
| Cubitt et al. (1998) | 3.1 vs. 3.3 | [redacted] | 0.685 | ✓ |
| Cubitt et al. (1998) | 3.2 vs. 3.3 | [redacted] | 0.120 | ✓ |
| Cox et al. (2014b) | PORpi vs. ImpureOT2 | [redacted] | 0.122 | ✓ |
| Cox et al. (2014b) | PORpi vs. ImpureOT3 | [redacted] | 0.988 | ✓ |
| Cox et al. (2014b) | PORpi vs. ImpureOT4 | [redacted] | 0.397 | ✓ |

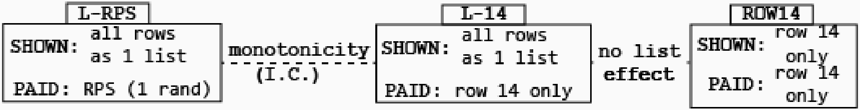
[redacted]

# Brown & Healy (2018)

| 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

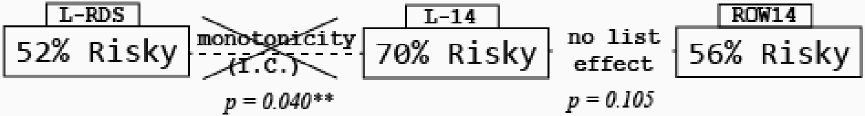
# Our Design




- Andreoni-Sprenger formatting
- Standard Ohio State subject pool.
- Between-subjects.
- Computerized.
  - List format: rows must be answered sequentially.
- Physical randomizing devices (die, bingo cage)
- No other tasks in the experiment.
- 60–63 subjects per treatment.
- Question: Do Row 14 choices differ by treatment?

# The Results

Row 14:




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

# Hypothesis

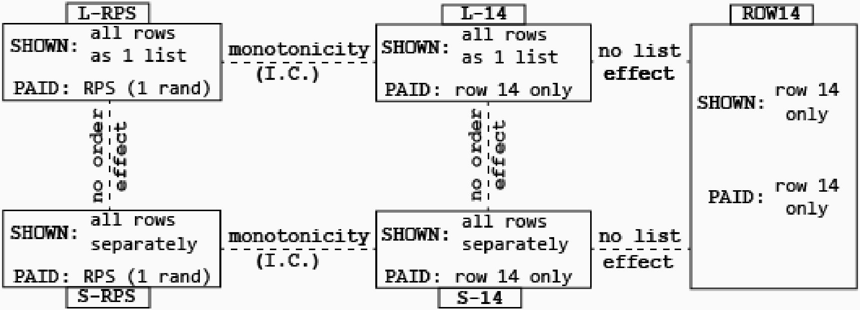
- Subjects are combining the decisions in a reduction-like way. *E.g.*: ‘When to switch?’.
- The ‘combining’ can be broken by separating the decisions.

## New Treatments

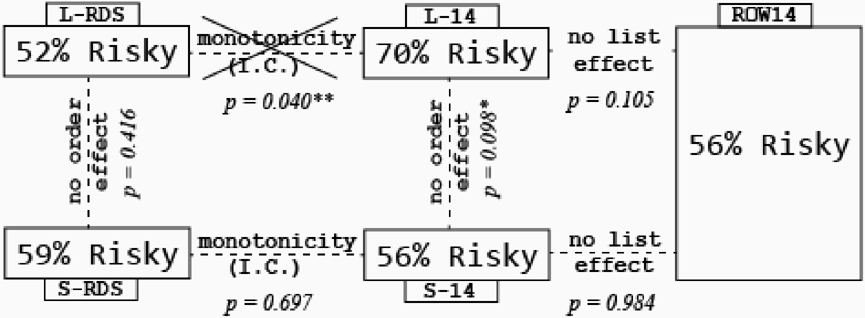
'Separated' treatments.

- Same 20 rows.
- Each shown on separate screen.
- Order of rows randomized for each subject.
- Still comparing RPS to Pay-14-Only.
- Still must answer every row, in order given.
  - First attempt: on paper. They shirked.
  - Second attempt: computerized, forced answers
- 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

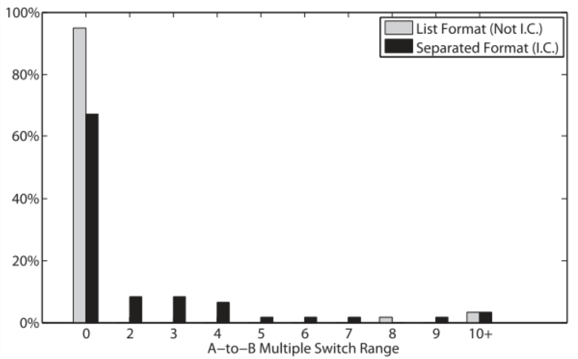
# The Results



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

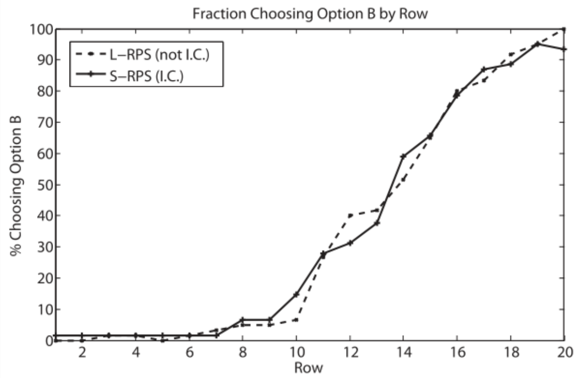
# The Cost of Separation

B-to-A (Risky-to-Safe) switches violate FOSD:
$Risky_{15}$ dominates $Risky_{14}$, but $Risky_{14} \succ Safe \succ Risky_{15}$




**LESSON: Separating decisions hurts consistency? NO!**
**The list format generates *false consistency*!**

# Biases Cancel Out




L-RPS was fine because “list effect” and “IC failure” canceled out!
I wouldn’t expect that to be true generally…

# Past Experiments

**Table 9**
Existing tests of incentive compatibility of the RPS mechanism that have no framing confounds. We describe each of these comparisons in the text below.

| Paper | Names of treatments | Presentation format | *p*-Value | RPS is I.C.? |
|---|---|---|---|---|
| Starmer and Sugden (1991) | A vs. B | List | 0.356 | ✓ |
| Starmer and Sugden (1991) | C vs. D | List | **0.043** | × |
| This paper | L-RPS vs. L-14 | List | **0.041** | × |
| This paper | S-RPS vs. S-14 | Separated | 0.697 | ✓ |
| Cubitt et al. (1998) | 3.1 vs. 3.3 | Separated | 0.685 | ✓ |
| Cubitt et al. (1998) | 3.2 vs. 3.3 | Separated | 0.120 | ✓ |
| Cox et al. (2014b) | PORpi vs. ImpureOT2 | Separated[a] | 0.122 | ✓ |
| Cox et al. (2014b) | PORpi vs. ImpureOT3 | Separated[a] | 0.988 | ✓ |
| Cox et al. (2014b) | PORpi vs. ImpureOT4 | Separated[a] | 0.397 | ✓ |

[a] Cox et al. (2014b) give subjects the choices on separate slips of paper, but the subjects could have arranged them into a list-like format if they wanted.

# Other Discussion of Separation

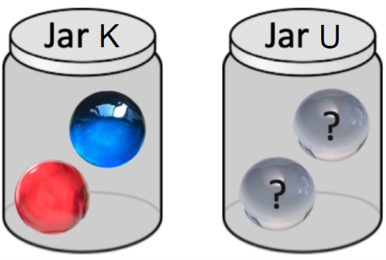
1. Kirby & Marakovic (1996) and Kirby et al. (1999)
    - Use scrambled lists in a field setting, including heroin addicts
2. Eckel et al. (2005)
    - Use scrambled with working poor
    - "we now believe that scrambling is a bad idea because it results in greater inconsistency and variance of responses."

## RPS for Hedging Ambiguity?

Is RPS used to hedge ambiguity?

- Oechssler Rau & Roomets (2019): No
    - Issues with their design
- Baillon Halevy & Li (2022)...

# 2-Urn Ellsberg Paradox




| | |
|---|---|
| $D_1$: | K $=$ \$2.00 if red from K<br>U $=$ \$2.10 if red from U |
| $D_2$: | K $=$ \$2.00 if blue from K<br>U $=$ \$2.10 if blue from U |

One paid randomly via coin flip

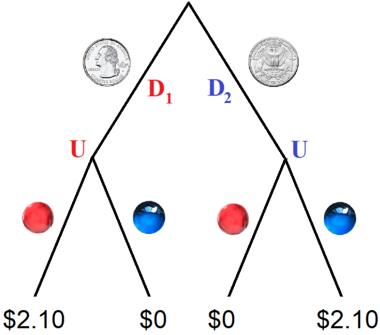
**Pr(Red in U)$\approx$Pr(Blue in U) & Ambiguity Averse:** K $\succ$ U and K $\succ$ U.

**Raiffa (1961):** Picking UU "hedges away" the ambiguity! UU $\succ$ KK

# How Hedging Works (Raiffa 1961)

Picking UU:

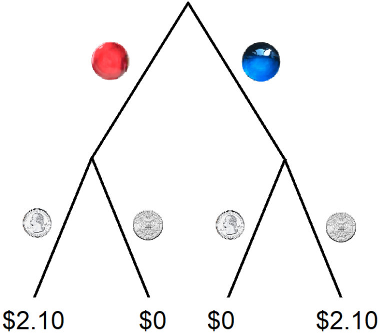
Original:




Ambiguity:

KK $\succ$ UU

Order-Reversed:



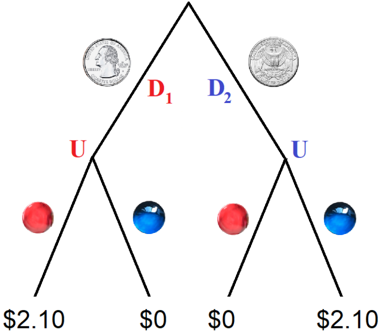

50-50 Lottery For Sure

UU $\succ$ KK

# How Hedging Works (Raiffa 1961)

Picking UU:

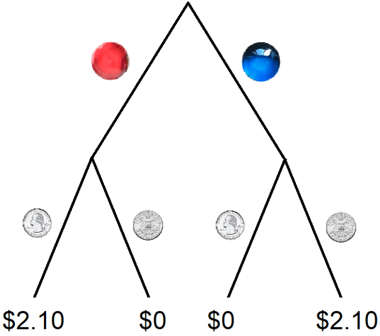
Original:




Ambiguity:

KK $\succ$ UU

Order-Reversed:




50-50 Lottery For Sure

UU $\succ$ KK

**Assumption:** Order Reversal

# Past Experiments

Order reversal has support...

- Coin before $\sim$ Coin after
  - Oechssler, Rau & Roomets (2019; ORR19)
  - Baillon, Halevy & Li (2022)

...yet people don't seem to appreciate hedging:

- Raiffa (1961), Dominiak & Schnedler (2011)
  - Ambiguity averse subjects don't value UU more than U and U
- ORR19 find mixed evidence for hedging
  - Amb. Averse & Pr(blue)$\approx$Pr(red) Subjects:
    - 50% consistent with hedging (or randomization)
    - Issues: Indifference & Cross-task contamination

# Past Experiments

Baillon, Halevy & Li (2022) (BHL22):

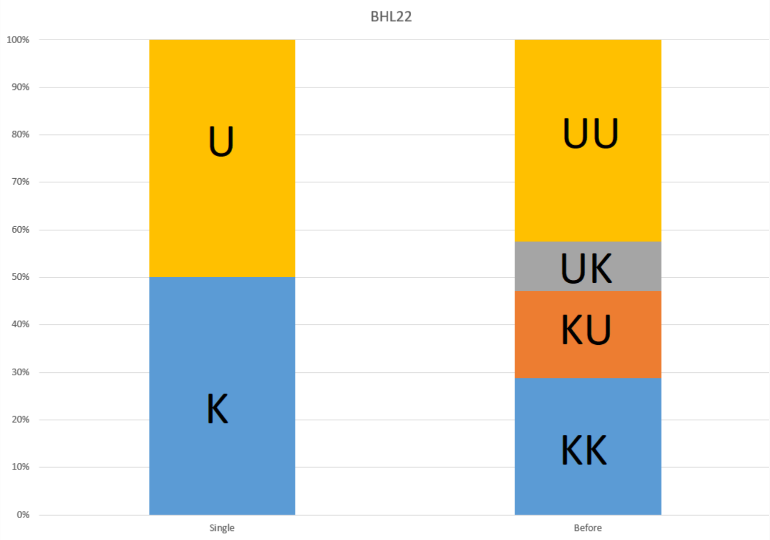
- "Single" Treatment:
  - $D_0 = \{K, K, U, U\}$
  - U or U $\Rightarrow$ Ambiguity neutral/loving *or* Pr(red)$\gtrless$Pr(blue)
  - K or K $\Rightarrow$ Strictly ambiguity averse *and* Pr(red)$\approx$Pr(blue)
  - 50% choose K or K

    $\Rightarrow$ 50% are Amb. Averse *and* Pr(red)$\approx$Pr(blue)

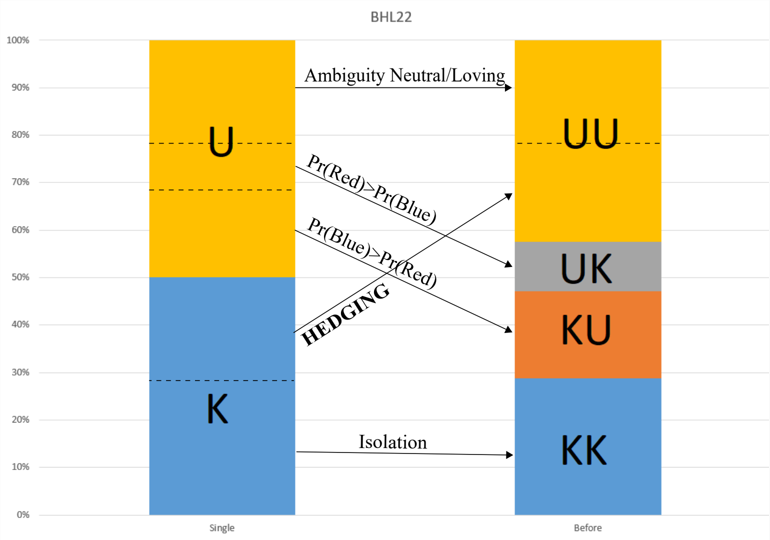
    this is a *lower bound* on Amb. Aversion
- "Before" Treatment:
  - $D_1 = \{K, U\}$, $D_2 = \{K, U\}$, coin flip **first**
  - What will Amb. Averse subjects pick?
    - Order Reversal + Hedging $\Rightarrow$ UU
    - "Isolation" $\Rightarrow$ KK
    - Pr(red)$\gtrless$Pr(blue) $\Rightarrow$ UK or KU

      (uses Azrieli et al. 2018, ignoring stochastic choice)

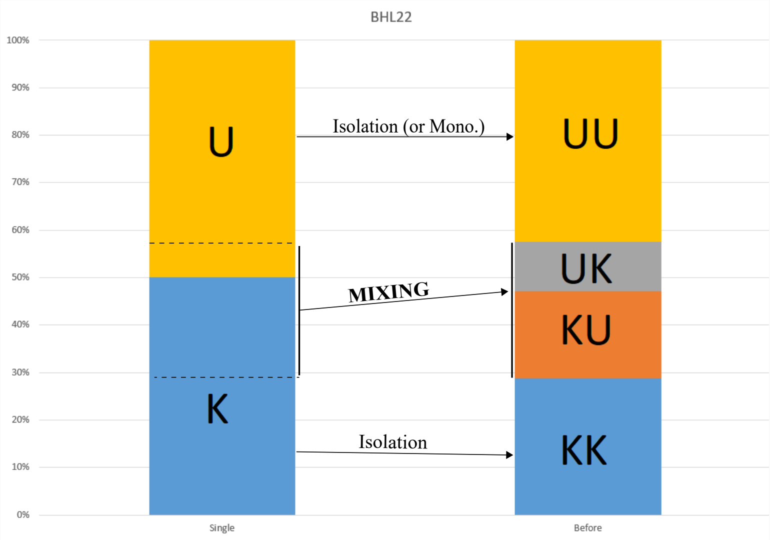
# BHL22: Results



# BHL22: Story 1



# BHL22: Story 2




Is it necessarily hedging?

# Pay All

Susan Laury's paper...

## Summary

- Theory: RPS generally fine *unless* subjects "reduce" (treat the experiment as one large decision)
- List format seems to encourage reduction, IC violations
- Separated format breaks reduction, restores IC
    - Separated *and* random order. Haven't tested which.
- List format generates *false consistency*
- Ambiguity:
    - RPS is not IC!
    - But is it really hedging??