Describing Deferred Acceptance and Strategyproofness to Participants: Experimental Analysis

Yannai A. Gonczarowski[1] Ori Heffetz[2] Guy Ishai[3] **Clayton Thomas**[4]

[1]Harvard [2]Hebrew University & Cornell [3]Hebrew University [4]Microsoft Research

Main Questions: How well do participants understand Deferred Acceptance (DA), and the strategyproofness (SP) property? Can changing descriptions improve understanding?

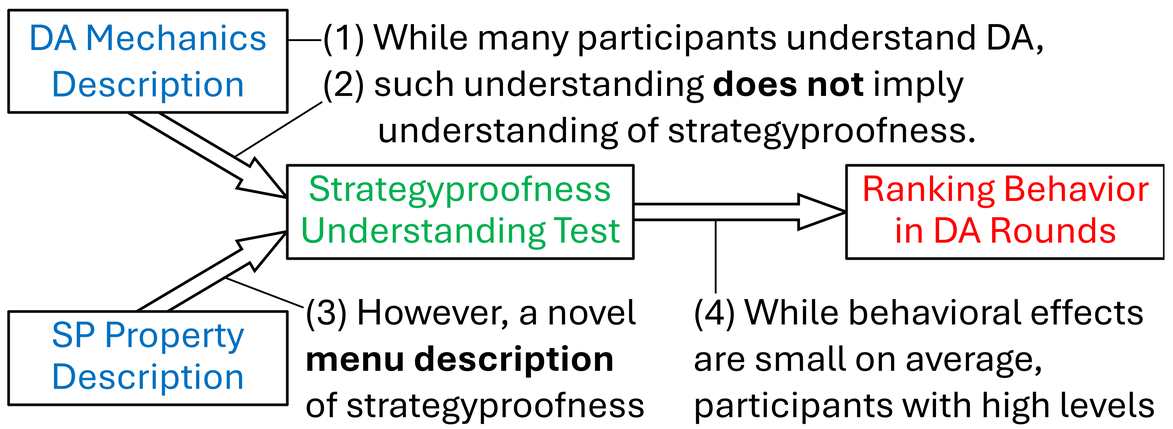
In an incentivized lab experiment, we describe to participants either:

(DA) The DA matching mechanism
- How to "mechanically" calculate DA's outcome (DA Mechanics)

(SP) The strategyproofness property
- Namely, $(\text{i's match})(\succ_i, \succ_{-i}) \succeq_i (\text{i's match})(\succ'_i, \succ_{-i})$ (SP Property)

We describe (DA) or (SP) to participants (in either classical or novel **menu** versions), measure strategyproofness understanding via specially-designed tests, and track effects of participants' ranking behavior.

Main Findings



Experiment Flow

After relaying the basic environment, we **describe DA, or SP,** to participants using one of five different descriptions.

We **train** participants on these descriptions, then they **play 10 rounds of DA**, then we **measure SP understanding** using novel tests.

How Do We Describe DA and Strategyproofness?

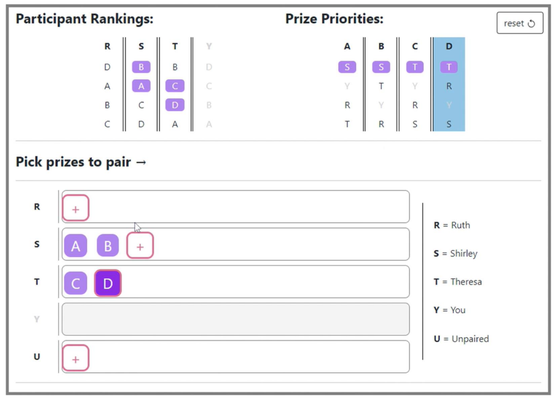
- We relay classic participant-proposing DA alg., and the textbook def'n of SP
- **And** we relay novel **menu** versions of each (Hammond, 1979)
- A **menu description** proceeds in two steps:
  (a) A "menu" of Obtainable Prizes, which *you* may receive, is determined using only the *other* participants' rankings.
  (b) Out of this menu, you receive your highest-ranked prize.

→ **SP easier to see**, in that it follows from a one-sentence proof! (Gonczarowski, Heffetz, & Thomas 2023; Katuščák & Kittsteiner 2022)

⇒ Our 5 Treatments (aka, 5 Descriptions of DA and SP)

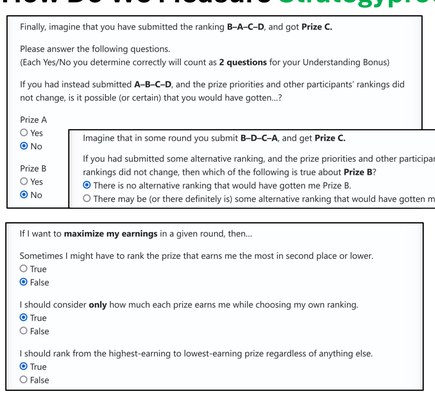
- Two **DA Mechanics** treatments: How one can "mechanically" calculate DA
  1. Traditional DA Mechanics (Trad-DA): Participant-proposing DA algorithm
  2. Menu DA Mechanics (Menu-DA): Test the menu description above, with **an explicit detailed algorithm** calculating the menu in step (a) (Gonczarowski, Heffetz, & Thomas 2023)
- Two **SP Property** treatments: Tell participants (only) that mechanism is SP
  3. Menu SP Property (Menu-SP): Test the menu description above, with **no details provided** about step (a)
  4. Textbook SP Property (Textbook-SP): An ordinary-language adaptation of $(\text{i's match})(\succ_i, \succ_{-i}) \succeq_i (\text{i's match})(\succ'_i, \succ_{-i})$
- Finally:
  5. Null: A (nearly) zero-information benchmark, where we tell participants almost nothing about how their outcome is calculated

How Do We Train Participants On Our Descriptions?



For DA Mechanics: Detailed training modules where participants use a GUI to calculate their DA outcomes in specific instance. (Menu-DA version shown here.)

How Do We Measure Strategyproofness Understanding?



Questions on applying the definition of strategyproofness in novel scenarios

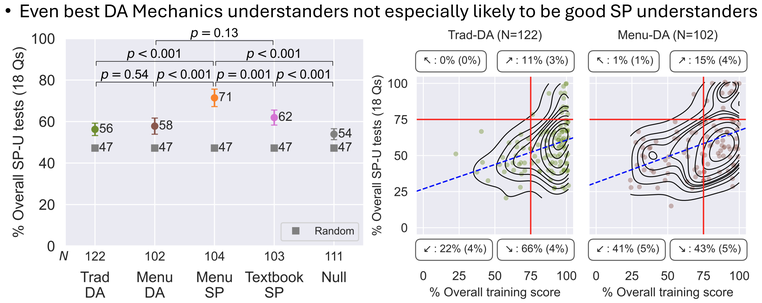
Questions on how participants can maximize earnings based on different DA attributes

(1) Many Participants Understand DA

- E.g., in Trad-DA, 81% of participants calculate their match correctly in ≥ 1 of our hardest, non-hand-held training questions; in Menu-DA this is 68%

(2) But, understanding DA *doesn't* imply understanding SP

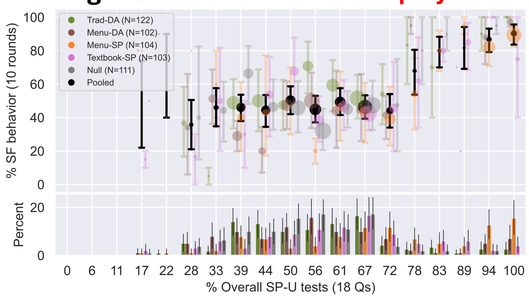
- In (both) DA Mechanics treatments, SP understanding scores **near Null treatment**
- Even best DA Mechanics understanders not especially likely to be good SP understanders



(3) Menu-SP conveys SP significantly better than others

- Raises participants' mean overall score, and sub-scores regarding definition of SP and how to maximize earnings.

(4) While effects on behavior are small on average, those with high levels of SP-U indeed play SF at very high rates



- **SP-U**: SP-understanding
- **SF**: straightforward ranking
- Mean SF rates:
  - 56% in Trad-DA; 50% in Menu-DA
  - 59% in Menu-SP; 53% in Textbook-SP
  - 48% in Null
  - Each with near-uniform spread
- Thus, participants (to some extent) act on SP-understanding
- Strongest effect in Menu-SP!
  - 32% of Menu-SP participants above 75% in both SP-U and SF; next-highest is 17% in Textbook-SP