Decision-Relevance and Irrelevant Questions: Deviation in Strategy-Proof Matching Mechanisms

The 18th Meeting of the Society for Social Choice and Welfare

Yuxing Liang
Concordia University

June 13, 2026

The Puzzle: Strategy-Proofness vs. Observed Behavior

Centralized matching markets (school choice, residency match) use **strategy-proof (SP)** mechanisms. Under SP, truthful reporting is a weakly dominant strategy:

$$\varphi_i(\succ_i, \succ_{-i}) \ \succeq_i \ \varphi_i(\widehat{\succ}_i, \succ_{-i}) \quad \text{for all } \widehat{\succ}_i, \succ_{-i} .$$

The Puzzle: Strategy-Proofness vs. Observed Behavior

Centralized matching markets (school choice, residency match) use **strategy-proof (SP)** mechanisms. Under SP, truthful reporting is a weakly dominant strategy:

$$\varphi_i(\succ_i, \succ_{-i}) \;\succeq_i\; \varphi_i(\widehat{\succ}_i, \succ_{-i}) \quad \text{for all } \widehat{\succ}_i,\, \succ_{-i}\,.$$

Yet agents routinely deviate from truthful reporting: Hassidim–Romm–Shorrer (*AER* 2021): large "obvious misrepresentation" rates. Rees-Jones (*GEB* 2018): suboptimal reports in the residency match.

Standard explanation

Agent failure. Cognitive limits prevent agents from recognising the dominant strategy. Prescription: educate agents, or redesign mechanism (Li *AER* 2017: OSP).

The Puzzle: Strategy-Proofness vs. Observed Behavior

Centralized matching markets (school choice, residency match) use **strategy-proof (SP)** mechanisms. Under SP, truthful reporting is a weakly dominant strategy:

$$\varphi_i(\succ_i, \succ_{-i}) \;\succeq_i\; \varphi_i(\widehat{\succ}_i, \succ_{-i}) \quad \text{for all } \widehat{\succ}_i, \succ_{-i}\,.$$

Yet agents routinely deviate from truthful reporting: Hassidim–Romm–Shorrer (*AER* 2021): large "obvious misrepresentation" rates. Rees-Jones (*GEB* 2018): suboptimal reports in the residency match.

Standard explanation

Agent failure. Cognitive limits prevent agents from recognising the dominant strategy. Prescription: educate agents, or redesign mechanism (Li *AER* 2017: OSP).

This paper

Mechanism over-elicits. A complete rank-order list encodes far more information than the allocation can use. Many "deviations" carry *zero* allocative consequence—they are rational under any positive deliberation cost.

A Critical Distinction

Two things often confused in the literature

Formal manipulation: a report that *strictly improves* the agent's allocation over truthful reporting. SP rules this out by definition.

Deviation from truth: the submitted ROL differs from the agent's true preference, *regardless* of whether it changes the allocation. **SP is silent on this.**

A Critical Distinction

Two things often confused in the literature

Formal manipulation: a report that *strictly improves* the agent's allocation over truthful reporting. SP rules this out by definition.

Deviation from truth: the submitted ROL differs from the agent's true preference, *regardless* of whether it changes the allocation. **SP is silent on this.**

This paper's question

How much of observed *deviation from truth* is also *allocatively inconsequential*? If the answer is "a lot," the prevailing cognitive-failure narrative is incomplete.

We formalise this by identifying which parts of an agent's rank-order list can affect her allocation, and which cannot.

Three Threads

1. **Theory.** Define *decision-relevance*: which report changes can affect the agent's allocation? Characterise the *truth-equivalence class* under SP, and prove it decomposes the report space cleanly. Apply to Serial Dictatorship and Deferred Acceptance.

2. **Empirics.** Use experimental data (Li *AER* 2017; Dreyfuss–Heffetz–Rabin *AEJ:Micro* 2022) to separate allocatively costless deviations (Type I) from costly ones (Type II). Test whether cognitive ability predicts the pattern.

3. **Design.** Interpret list-length limits and sequential elicitation as tools that reduce or eliminate zero-incentive questions. Calibrate truncation length against two accuracy criteria.

Core message: a large share of observed "manipulation" under SP is not manipulation in any formal sense—it is the mechanism's own informational redundancy.

The Basic Observation: Two Views of a Rank-Order List

A submitted ROL over m objects can be read in two ways simultaneously.

Agent's view (elicitation)

Forming a strict ranking requires $m-1$ **adjacent decisions**: is o_1 better than o_2? Is o_2 better than o_3? ...

The remaining $\binom{m}{2} - (m-1)$ pairwise comparisons are **logically implied by transitivity**—the agent need not decide them separately.

The Basic Observation: Two Views of a Rank-Order List

A submitted ROL over m objects can be read in two ways simultaneously.

Agent's view (elicitation)

Forming a strict ranking requires $m-1$ **adjacent decisions**: is o_1 better than o_2? Is o_2 better than o_3? ...
The remaining $\binom{m}{2} - (m-1)$ pairwise comparisons are **logically implied by transitivity**—the agent need not decide them separately.

Mechanism's view (incentives)

A submitted ROL **certifies** $\binom{m}{2}$ pairwise answers—for any pair $\{a, b\}$, the list says who comes first.
The mechanism can in principle condition on any permutation, so it "sees" $\binom{m}{2}$ inputs, whether or not the agent deliberated over each one.

m	**Certified pairs** $\binom{m}{2}$	**Adj. decisions** $m-1$	Possible outcomes	Pair ratio
10	45	9	11	4×
100	4,950	99	101	49×
500	124,750	499	501	249×

Two distinct redundancy questions arise—one for each view: Which certified pairs does SP require to be allocatively irrelevant? (*mechanism view, ratio* r^{pos}) Which adjacent decisions does SP require to be wasted? (*agent view, ratio* r^{adj})
Both have precise answers. We develop them for SD and DA.

Model and Decision-Relevance

$N = \{1, \ldots, n\}$ agents, $O = \{o_1, \ldots, o_m\}$ objects. Each agent i has strict preference $\succ_i$ over $O \cup \{\emptyset\}$. A mechanism $\varphi : \prod_j \mathcal{P}_j \to \mathcal{M}$ maps reported profiles to assignments.

Definition: decision-relevance of a deviation

Fix mechanism φ, agent i, true preference $\succ_i$, and a feasible deviation $T : \mathcal{P}_i \to \mathcal{P}_i$.

T is **decision-relevant** at $\succ_i$ if $\exists\ \succ_{-i}$ such that $\varphi_i(\succ_i, \succ_{-i}) \neq \varphi_i(T \succ_i, \succ_{-i})$.

T is **decision-irrelevant** if $\varphi_i(\succ_i, \succ_{-i}) = \varphi_i(T \succ_i, \succ_{-i})$ for *every* $\succ_{-i}$.

Model and Decision-Relevance

$N = \{1, \ldots, n\}$ agents, $O = \{o_1, \ldots, o_m\}$ objects. Each agent i has strict preference $\succ_i$ over $O \cup \{\emptyset\}$. A mechanism $\varphi : \prod_j \mathcal{P}_j \to \mathcal{M}$ maps reported profiles to assignments.

Definition: decision-relevance of a deviation

Fix mechanism φ, agent i, true preference $\succ_i$, and a feasible deviation $T : \mathcal{P}_i \to \mathcal{P}_i$.

T is **decision-relevant** at $\succ_i$ if $\exists\ \succ_{-i}$ such that $\varphi_i(\succ_i, \succ_{-i}) \neq \varphi_i(T \succ_i, \succ_{-i})$.
T is **decision-irrelevant** if $\varphi_i(\succ_i, \succ_{-i}) = \varphi_i(T \succ_i, \succ_{-i})$ for *every* $\succ_{-i}$.

Two natural primitives for T:

Adjacent swap A_r

Swaps objects at positions r and $r+1$. Exactly one local ranking margin changes. There are $m-1$ such margins.

$$r_i^{adj} = \frac{|\{r : A_r \text{ relevant}\}|}{m-1}$$

Positional swap P_{ab}

Exchanges positions of a and b. For non-adjacent a, b: changes $2(s-r)-1$ comparisons simultaneously—not just the direct a–b comparison.

$$r_i^{pos} = \frac{|\{\{a, b\} : P_{ab} \text{ relevant}\}|}{\binom{m}{2}}$$

The Report-Space Decomposition

For a fixed SP mechanism, two different ROLs may produce the same allocation for agent i against *every* profile of others' reports.

Definition: truth-equivalence class

$\widehat{\succ}_i \sim_{\varphi,i} \succ_i$ if $\varphi_i(\widehat{\succ}_i, \succ_{-i}) = \varphi_i(\succ_i, \succ_{-i})$ for all $\succ_{-i}$. The **truth-equivalence class** is $[\succ_i]_{\varphi,i} = \{\widehat{\succ}_i : \widehat{\succ}_i \sim_{\varphi,i} \succ_i\}$.

The Report-Space Decomposition

For a fixed SP mechanism, two different ROLs may produce the same allocation for agent i against *every* profile of others' reports.

Definition: truth-equivalence class

$\widehat{\succ}_i \sim_{\varphi,i} \succ_i$ if $\varphi_i(\widehat{\succ}_i, \succ_{-i}) = \varphi_i(\succ_i, \succ_{-i})$ for all $\succ_{-i}$. The **truth-equivalence class** is $[\succ_i]_{\varphi,i} = \{\widehat{\succ}_i : \widehat{\succ}_i \sim_{\varphi,i} \succ_i\}$.

Proposition (Report-Space Decomposition under SP)

If φ is strategy-proof, then every alternative report falls into exactly one of two regions:

1. **Zero-incentive region** $[\succ_i]_{\varphi,i}$: assignment-identical to truth against every $\succ_{-i}$; expected incentive loss $= 0$ under *any* prior and utility.
2. **Positive-cost region** $\mathcal{P}_i \setminus [\succ_i]_{\varphi,i}$: truth weakly dominates it at every $\succ_{-i}$; at some witness profile truth is strictly better.

Key implication. Any deviation inside $[\succ_i]_{\varphi,i}$ is not a formal manipulation—it is a report with zero expected cost, rational under any positive deliberation cost.

Report Distance Is Not Incentive Loss

Empirical work records whether submitted lists differ from truth. The theory asks: *does the difference change the assignment?*

Example. Serial Dictatorship, agent i first in queue, true preference $a \succ_i b \succ_i c \succ_i d$.

Large distance, zero incentive loss

Report $\widehat{\succ}_i^1$: $a \succ d \succ c \succ b$.
Keeps a at top; reverses bottom block. Kendall distance: $d_R = 3$.
Agent i still receives a against every others' profile.

$$\widehat{\succ}_i^1 \in [\succ_i]_{\varphi,i}, \quad \ell_i(\widehat{\succ}_i^1) = 0.$$

Small distance, positive loss

Report $\widehat{\succ}_i^2$: $b \succ a \succ c \succ d$.
One adjacent swap; Kendall distance $d_R = 1$.
Assignment shifts from a to b. Strictly worse whenever $u_i(a) > u_i(b)$.

$$\widehat{\succ}_i^2 \notin [\succ_i]_{\varphi,i}, \quad \ell_i(\widehat{\succ}_i^2) > 0.$$

Measured "manipulation rates" conflate these two cases. Our framework separates them.

Serial Dictatorship: Structure of the Zero-Incentive Region

Serial Dictatorship (SD): agents pick in a fixed queue; each takes her best available object.

Key observation. Agent i at queue position k receives one of her *top-k* objects—at most $k-1$ are taken by those ahead.

Position $k = 1$ (first)

Only top choice matters. ROL below rank 1 is irrelevant. Zero-incentive region: $(m-1)!$ reports.

Position $k = 2$

Top-2 ranking matters. ROL below rank 2 is irrelevant. Zero-incentive region: $(m-2)!$ reports.

Position $k = m$ (last)

Receives whatever remains. *Entire* ROL is irrelevant. Zero-incentive region: $m!$ reports.

Proposition: exact truth-equivalence class under SD

If $k < m$, then $\widehat{\succ}_i \in [\succ_i]_{SD,i}$ iff $\widehat{\succ}_i$ ranks $o_1, \ldots, o_k$ first and in the truthful order; objects $o_{k+1}, \ldots, o_m$ may be ordered arbitrarily. Hence $|[\succ_i]_{SD,i}| = (m-k)!$.

SD Example: $n = m = 4$

True preference $a \succ_i b \succ_i c \succ_i d$. Six pairwise queries: $q_{ab}, q_{ac}, q_{ad}, q_{bc}, q_{bd}, q_{cd}$.

Position k	Relevant pair queries	Irrelevant pair queries	r_i^{pos}	r_i^{adj}	Equiv. ROLs
1	$q_{ab},\ q_{ac},\ q_{ad}$	$q_{bc},\ q_{bd},\ q_{cd}$	$1/2$	$1/3$	$3! = 6$
2	$q_{ab},\ q_{ac},\ q_{ad},\ q_{bc},\ q_{bd}$	q_{cd}	$5/6$	$2/3$	$2! = 2$
3	all 6	—	1	1	$1! = 1$
4	—	all 6	0	0	$4! = 24$

Block structure under SD. Irrelevant queries involve only objects ranked below position k. Any permutation of the bottom $m - k$ objects leaves the assignment unchanged.

The agent at $k = 1$ can submit $3! = 6$ different ROLs that all produce the same outcome. These are not manipulations—they are answers to questions the mechanism does not use.

Random Serial Dictatorship: Quantifying the Waste

Under **RSD**, queue position k is drawn uniformly at random. Averaging the relevance ratios over all positions:

Full (positional-pair) ratio

$$\bar{r}^{pos}(\mathrm{RSD}) = \frac{2m-1}{3m} \xrightarrow{m\to\infty} \frac{2}{3}$$

$\Rightarrow$ **One-third** of encoded pairwise comparisons are irrelevant on average.

Adjacent ratio

$$\bar{r}^{adj}(\mathrm{RSD}) = \frac{1}{2}$$

$\Rightarrow$ **One-half** of the agent's independent ranking decisions are wasted on average.

Why two ratios? r^{pos} measures information the *mechanism* wastes (mechanism-side view). r^{adj} measures deliberation the *agent* wastes (agent-side view). The gap reflects logical redundancy from transitivity: non-adjacent pairs are encoded in the ROL but cost nothing extra once the adjacent ordering is decided.

Even at its simplest, complete-ROL elicitation must waste a constant fraction of the agent's effort.

Deferred Acceptance: Why the Structure Changes

DA is used in NYC, Boston, Chicago school choice and the National Resident Matching Program.

Student-proposing DA: students propose in the order of their submitted ROL; schools tentatively hold the highest-priority applicants and reject the rest; rejected students propose to their next choice; repeat until stable.

Two features that complicate the relevance structure:

1. Proposal order matters

The agent proposes to schools *in the order* of her submitted ROL. Swapping a and b changes which of $\{a, b\}$ she proposes to first.
This is the **direct pathway**: swap may change allocation between a and b.

2. Rejection chains exist

When i displaces j from school s, j proposes elsewhere, possibly displacing k, and so on.
This is the **rejection-chain pathway**: swap of a and b may change allocation to some $c \notin \{a, b\}$.
This has no analogue in SD.

Because of rejection chains, DA's zero-incentive region is harder to characterise—and likely much larger in large markets.

DA in Large Markets: Ex-Ante Realized Use

Setup: balanced random DA; $n = m$, unit capacities, uniform independent preferences and priorities.

Theorem (Pittel 1989): typical proposal reach is $\Theta(\log m)$

In the balanced random DA market, a typical agent is matched to one of her top-$\Theta(\log m)$ choices: $\mathbb{E}[\rho_i^{full}] = \Theta(\log m)$. Objects ranked below the **proposal reach** are never proposed to.

Two distinct concepts: realized path vs. counterfactual relevance

ρ_i^{full} is a *probabilistic realized-path* measure. A non-reached object could still be decision-relevant in the worst-case (global) sense—under different priorities, it might matter. These are separate notions; the Pittel result gives the *ex-ante* version.

Adjacent realized-use

$$u_i^{adj} = \frac{\rho_i^{full} - 1}{m - 1} = \Theta\Big(\frac{\log m}{m}\Big) \to 0.$$

Pairwise exposure

$$u_i^{pair} = \frac{\binom{m}{2} - \binom{m-\rho_i^{full}}{2}}{\binom{m}{2}} = \Theta\Big(\frac{\log m}{m}\Big) \to 0.$$

DA: Scale of Unused List and Who Deviates More

Quantitative size of the unreached tail:

m	Avg reach	Adj. unreached	Pairs unexposed
50	≈ 3.9	$\approx 94\%$	$\approx 85\%$
100	≈ 4.6	$\approx 96\%$	$\approx 91\%$
500	≈ 6.2	$\approx 99\%$	$\approx 98\%$
1000	≈ 6.9	$\approx 99\%$	$\approx 99\%$

Unreached fraction $\to 1$ as $m \to \infty$; DA's informational waste is quadratically worse than SD's in large markets.

DA: Scale of Unused List and Who Deviates More

Quantitative size of the unreached tail:

m	Avg reach	Adj. unreached	Pairs unexposed
50	≈ 3.9	$\approx 94\%$	$\approx 85\%$
100	≈ 4.6	$\approx 96\%$	$\approx 91\%$
500	≈ 6.2	$\approx 99\%$	$\approx 98\%$
1000	≈ 6.9	$\approx 99\%$	$\approx 99\%$

Unreached fraction $\to 1$ as $m \to \infty$; DA's informational waste is quadratically worse than SD's in large markets.

Heterogeneity: priority concentration predicts deviation

Concentrated priority

Strong priority at a small set of preferred schools. Outcome is predictable; a short prefix often determines the match. $\Rightarrow$ Large fraction of lower-list questions are likely zero-incentive.

Diffuse priority

Moderate priority across many schools. Outcome depends on many margins and rejection-chain contingencies. $\Rightarrow$ More of the ROL may matter.

Empirical prediction

Measured deviation rates should be **higher** for agents with concentrated priority—not because they are less rational, but because the mechanism over-asks them more.

Does the Framework Predict Observed Behavior?

The underlying experimental data are from Li (AER 2017). We also use the reanalysis of these data by Dreyfuss, Heffetz, and Rabin (2022).

Setup: 4 agents, 4 prizes, RSD; subjects submit complete ROLs; induced preferences observed. Both **one-shot** and **multi-round** (8 periods) treatments.

Our classification:

- **Type I deviation:** a submitted ROL that differs from truth but lies *inside* $[\succ_i]_{\varphi,i}$—allocatively equivalent to truth. No loss is possible.
- **Type II deviation:** a submitted ROL *outside* $[\succ_i]_{\varphi,i}$—truth would have been weakly better.

Does the Framework Predict Observed Behavior?

The underlying experimental data are from Li (AER 2017). We also use the reanalysis of these data by Dreyfuss, Heffetz, and Rabin (2022).

Setup: 4 agents, 4 prizes, RSD; subjects submit complete ROLs; induced preferences observed. Both **one-shot** and **multi-round** (8 periods) treatments.

Our classification:

- **Type I deviation:** a submitted ROL that differs from truth but lies *inside* $[\succ_i]_{\varphi,i}$—allocatively equivalent to truth. No loss is possible.
- **Type II deviation:** a submitted ROL *outside* $[\succ_i]_{\varphi,i}$—truth would have been weakly better.

Three questions:

1. Do deviations concentrate on theoretically irrelevant comparisons?
2. What share of observed "manipulation" is Type I (allocatively costless)?
3. Is the pattern explained by cognitive ability, or by the incentive structure?

Main Empirical Findings

Finding 1: Pair-level deviation rates (multi-round data)

Pair type	Deviation rate	N (pairs)
Decision-relevant	12.70%	2,520
Decision-irrelevant	**17.89%**	1,800

Deviations are **41% more frequent** on irrelevant comparisons. Interaction significant under multiple SE clustering strategies (see appendix).

Main Empirical Findings

Finding 1: Pair-level deviation rates (multi-round data)

Pair type	Deviation rate	N (pairs)
Decision-relevant	12.70%	2,520
Decision-irrelevant	**17.89%**	1,800

Deviations are **41% more frequent** on irrelevant comparisons. Interaction significant under multiple SE clustering strategies (see appendix).

Finding 2: Round-level Type I classification (multi-round data)

	Misreporting rounds	Of which: pure Type I
Multi-round	209	**80 rounds (38.3%)**

38% of misreporting rounds involve *only* Type I deviations: the submitted ROL is allocatively identical to truth, so no strategic payoff is possible.

One-shot data: pattern weaker, suggesting learning matters.

Ruling Out Cognitive Confusion

Alternative hypothesis: bottom-only deviations reflect cognitive confusion that happens to fall in the irrelevant region by chance.

Test: if confusion drives the pattern, low-GPA subjects should deviate more on irrelevant comparisons than high-GPA subjects.

Group	Misreporting rounds	Type I share	
High GPA	116	38.79%	
Low GPA	93	37.63%	
Difference		**1.16 pp**	(statistically indistinguishable)

The cognitive-ability gap is essentially zero. Combined with the queue-position test (framework predicts $r^{pos} = 0$ at $k = m = 4$; data show 0% Type I at that position), a pure confusion story is ruled out.

Empirical summary

Deviations concentrate where theory predicts. 38% of misreporting rounds are allocatively costless. Cognitive ability does not predict the pattern. **The data are consistent with rational economising over a redundant elicitation interface.**

Design Response: Reducing Zero-Incentive Questions

If complete-ROL elicitation over-asks, two design responses are natural.

Truncation (static)

Require lists of length $K < m$.

In practice: NYC limits to 12 schools; Boston to 10.

Key question: how short can K be without changing the matching? Answer depends on *which accuracy criterion* is chosen.

Sequential elicitation (adaptive)

Ask each question only when its answer is immediately consequential.

Studied in: Bó–Hakimov (*JET* 2022; *EJ* 2023); related to Li's OSP.

Property: by construction, every elicited answer is decision-relevant. No zero-incentive questions.

Truncation is simple and static; sequential is adaptive but requires real-time interaction. Both shrink or eliminate the zero-incentive region.

How Short Can Lists Be? Two Criteria

Two accuracy standards

Individual accuracy: agent's proposal reach fits in her list with probability $\geq 1 - \delta$.
$K_{\text{ind}} \approx 3.2 \log m$.

Full-matching exactness: the entire match equals what full-list DA produces, w.p. $\geq 1 - \delta$.
$K_{\text{full}} \approx 1.45(\log m)^2$.

Simulation at $\delta = 0.05$, balanced random DA:

m	Avg reach	K_{ind}	K_{full}
50	4.2	12	27
100	5.0	14	37
500	6.8	20	65
1000	7.4	21	73

NYC calibration

NYC accepts up to **12 schools**; effective $m \approx 500$.

$K_{\text{ind}}(500, 0.05) = 20$
$K_{\text{full}}(500, 0.05) = 65$

NYC's 12-school limit is below even the individual-accuracy threshold under the idealized baseline.

Caveat

NYC has correlated preferences, capacities $q > 1$, screened admissions. Numbers are *illustrative*; the qualitative separation between criteria is robust.

The same fixed list length delivers very different welfare guarantees depending on which standard applies.

Summary

1. A clean conceptual distinction. Formal manipulation (profitable misreport, ruled out by SP) vs. deviation from truth (SP silent). Decision-relevance formalises which parts of a ROL can affect the allocation.

2. Theory: report-space decomposition. Under SP, every alternative report is in exactly one region: the zero-incentive class (allocatively equivalent to truth) or the positive-cost class (truth weakly dominates, strictly at some profile).

3. Quantification. SD/RSD: 1/3 of pairwise comparisons and 1/2 of adjacent decisions irrelevant on average. DA: realized reach is $\Theta(\log m)$; over 97% of submitted list unreached in large markets.

4. Empirics. 38% of misreporting rounds in Li's multi-round data are pure Type I (allocatively costless). Pattern is not explained by cognitive ability.

5. Design. $K_{\text{individual}} \sim \log m$ vs. $K_{\text{full}} \sim (\log m)^2$. Sequential elicitation eliminates zero-incentive questions by construction.

Observed deviations from truth are not always strategic failures.
Often they are the mechanism's own redundancy, reflected back as noise.

Open Questions

- **Theoretical:** Full characterisation of the truth-equivalence class under DA: when exactly does a non-reached object become decision-relevant? (Rejection-chain pathway; related to bossiness of DA, cf. Raghavan 2018.)
- **Empirical:** Test the concentration prediction on field-scale DA data (NYC, Boston, NRMP): does deviation rate increase with priority concentration, controlling for other factors?
- **Robustness:** Extension to correlated preferences, many-to-one matching ($q > 1$), screened admissions.
- **Design:** Welfare comparison of truncation vs. sequential implementation under heterogeneous deliberation costs.
- **Other mechanisms:** Structure of the zero-incentive region under Top Trading Cycles; generalisation beyond matching.

Thank You

"Observed deviations are the mechanism's own redundancy,
reflected back as noise in the data.

The resolution is not to educate the agents
but to reform the questions."

Yuxing Liang
Concordia University

yuxing.liang@concordia.ca

Appendix: Deriving the RSD Relevance Ratios

Under RSD each queue position $k \in \{1, \dots, m\}$ occurs with probability $1/m$.

Adjacent ratio

For $k < m$: relevant adjacent margins are positions $1, \dots, k$ (the first k gaps). For $k = m$: no adjacent margin is relevant (agent receives the unique remaining object).

$$\bar{r}^{adj}(\mathrm{RSD}) = \frac{1}{m} \sum_{k=1}^{m-1} \frac{k}{m-1} = \frac{1}{m(m-1)} \cdot \frac{m(m-1)}{2} = \frac{1}{2}.$$

Full positional ratio

At position $k < m$: a positional pair $\{a, b\}$ is relevant iff the higher-ranked object lies in the top k, so $|R_i^{pos}(k)| = \sum_{r=1}^{k}(m-r)$.
Averaging over positions,

$$\bar{r}^{pos}(\mathrm{RSD}) = \frac{1}{m} \sum_{k=1}^{m-1} \frac{\sum_{r=1}^{k}(m-r)}{\binom{m}{2}} = \frac{2m-1}{3m}.$$

Appendix: DA Two Pathways Through Which a Swap Affects Allocation

A swap of a and b in agent i's ranking can change her allocation through two distinct channels:

Direct pathway

Allocation changes **between** a **and** b themselves. Active when both a and b are in the agent's option set (Barberà 1983). Same logic as SD.

Rejection-chain pathway

Allocation changes to $c \notin \{a, b\}$ because the swap initiates a different rejection chain. No analogue in SD. Related to *bossiness* of DA (Raghavan 2018).

A query is decision-irrelevant only when *both* pathways are blocked: (a) $\{a, b\} \cap O_i = \emptyset$ (neither achievable); (b) no element of O_i lies between a and b in $\succ_i$; (c) the swap does not alter any rejection chain reaching an object in O_i.

Full characterisation of conditions (b)–(c) remains open. We provide sufficient conditions and use the realized-reach proxy for quantification.

Appendix: Identical Priorities $\Rightarrow$ DA $=$ SD

Special case: if all schools share the same priority order over students, student-proposing DA reduces to SD with that priority as the queue.

Intuition: the highest-priority student is held at her first proposal everywhere; she is never displaced. The next student's sequence is determined by what the first took. Inductively, each student picks the best available school—exactly SD.

Implication: the SD truth-equivalence class characterisation (bottom-$(m-k)!$ equivalences) is a special case of the DA class. With heterogeneous priorities, rejection chains enlarge the zero-incentive region beyond the SD bottom block—making DA's informational waste even more severe than the SD baseline.

Appendix: Robustness of Pair-Level Regression

Specification. $\text{Deviation}_{i,pair} = \alpha + \beta \cdot \text{Irrelevant} + \gamma \cdot \text{MultiRound} + \delta \cdot (\text{Irrelevant} \times \text{MultiRound}) + \varepsilon$

SE type	Interaction $\hat{\delta}$	p-value
HC1	0.087	0.000
Cluster by subject	0.087	0.026
Cluster by agent-round	0.087	0.024
Two-way cluster (subject $\times$ session)	0.087	0.064
With pair FE + controls (payoff gap, true rank, queue, adjacent)		
HC1	0.097	0.000
Cluster by subject	0.097	0.016

The interaction is positive and significant across all specifications. Controls and standard clustering do not overturn the finding. The repeated-setting strengthens the irrelevant-comparison channel.

Appendix: Key References I

Ashlagi, Kanoria & Leshno (*JPE*, 2017). Unbalanced random matching markets.

Bó & Hakimov (*JET*, 2022; *EJ*, 2023). Pick-an-object mechanisms; iterative deferred acceptance.

Caplin & Dean (*AER*, 2015). Revealed preference, rational inattention, costly information acquisition.

Carroll (*Econometrica*, 2012). When are local incentive constraints sufficient?

Che & Tercieux (*JPE*, 2019). Efficiency and stability in large matching markets.

Coles & Shorrer (*GEB*, 2014). Optimal truncation in matching markets.

Dreyfuss, Heffetz & Rabin (*AEJ:Micro*, 2022). Expectations-based loss aversion and seemingly dominated choices in SP mechanisms.

Ehlers (*MOR*, 2008). Truncation strategies in matching markets.

Hassidim, Romm & Shorrer (*AER*, 2021). The mechanism is truthful, why aren't you?

Kojima & Pathak (*AER*, 2009). Incentives and stability in large two-sided matching markets.

Li (*AER*, 2017). Obviously strategy-proof mechanisms.

Pittel (*SIAM J. Disc. Math.*, 1989). The average number of stable matchings.

Raghavan, M. (*WP*, 2018). Influence and bounded rationality in matching markets.

Rees-Jones (*GEB*, 2018). Suboptimal behavior in strategy-proof mechanisms.

Roth & Rothblum (*Econometrica*, 1999). Truncation strategies in matching markets.

Sato (*JET*, 2013). A sufficient condition for the equivalence of strategy-proofness and nonmanipulability by adjacent preferences.