Report-Dependent Utility and Strategy-Proofness

Vincent Meisner (TU Berlin)

Discussion Paper No. 289

October 21, 2021

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Georg Weizsäcker, Ph.D., Humboldt University Berlin, 10117 Berlin, Germany
info@rationality-and-competition.de

Report-dependent utility and strategy-proofness*

Vincent Meisner†

August 16, 2021

Abstract

Despite the truthful dominant strategy, participants in strategy-proof mechanisms submit manipulated preferences. In our model, participants dislike rejections and enjoy the confirmation from getting what they declared most desirable. Formally, the payoff from a match decreases in its position in the submitted ranking such that a strategic trade-off between preference intensity and match probability arises. This trade-off can trigger the commonly observed self-selection strategies. We show that misrepresentations can persist for arbitrarily small report-dependent components. However, honesty is guaranteed to be optimal if and only if there is no conflict between the quality and feasibility of a match.

JEL-Classification: D47, D78, D81, D91.
Keywords: Market design, matching, school choice, self-regarding preferences, strategy-proof mechanisms.

*I thank Inácio Bó, Rustamdjan Hakimov and Jonas von Wangenheim Financial support by Deutsche Forschungsgemeinschaft (CRC/TRR 190, project 280092119) and UniCredit Foundation (Modigliani Grant) is gratefully acknowledged.

†Technical University Berlin, Straße des 17. Juni 135, 10623 Berlin, Germany, vincent.meisner@tu-berlin.de.

1 Introduction

Since revealing the true preferences is a dominant strategy[1] in strategy-proof mechanisms, there is no gain from sophisticated strategizing or costly information acquisition about others. Consequently, such mechanisms are deemed fair: they "level the playing field." However, there is extensive experimental and field evidence (Hakimov and Kübler, 2021; Hassidim et al., 2017a) that participants misrepresent their preferences, in particular, by skipping popular options in the submitted ranking. To remedy possible negative effects on justified envy and efficiency, understanding what is behind this phenomenon is important. Instead of designating non-truthful strategies a mistake, researchers recently suggested more complex preferences under which such strategies can be optimal. To identify the origin of such deviations, testable predictions for all competing theories are needed.

In our model, report-dependent utility introduces a strategic motive into direct strategy-proof mechanisms. On top of the utility garnered from the assignment, a participant receives an additional payoff that decreases in the rank of the matched option in her submitted rank-ordered list (ROL). This component can be positive and, for instance, reflect the "warm glow" from being accepted at a top choice, or the enjoyment from telling other participants (and herself) that she did not have any rejections and "got exactly what she asked for." When she is assigned to an option ranked at the bottom, this utility can turn negative to reflect, for instance, the frustration from having been rejected by every higher-ranked option or the consternation that the reported preferences are not mutual. Striving for the former positive emotions or avoiding the latter negative feedback can upset the strategy-proofness and lead to the observed patterns of misrepresentations. Disregarding such emotional factors, report-dependent utility can also arise due to signaling motives in a larger game.[2]

One may think that report-dependent utility is negligibly small in real-life settings and, thus, its effect on reported preferences in strategy-proof mechanisms vanishes. However, for any ROL, we can construct a robust set of beliefs such that this ROL is strictly optimal for any report-dependent and report-independent preference. By Proposition 1, participants may strictly prefer non-truthful ROLs when arbitrarily small report-dependent utility is added to arbitrarily strong "standard preferences." For instance, the constructed beliefs are reasonable for low-priority participants, and we predict the pattern suggested by the data: such participants order options by chances of admission rather than preferences. Truthful reporting

[1] In line with much of the mechanism-design literature, we are sloppy in the use of the game-theoretic term "dominant" and employ it as a synonym for "always optimal," see Börgers (2015, Chapter 4) for a discussion.

[2] For instance, if one side's ROL is hard information, while the priorities of the other side are unknown, a match with a reported top choice can be used as information consistent with a high priority to a third party with similar preferences but less information than the other side. A proposer might also be interested in signaling to the receivers that her preferences are in line with¡ theirs.

is most prevalent when there is no conflict between preferences and admission probabilities. We confirm this observation in Proposition 2: even arbitrarily large report-dependent payoffs cannot render deviations from the truth profitable in such cases.

In their seminal experimental paper on school choice, Chen and Sönmez (2006) coin the small-school bias and district-school bias: participants hide their preferences for competitive options or fake a preference for options where they expect high chances of admission. A self-selection strategy can manifest itself in both biases. For instance, Chen and Pereyra (2019) link Mexican school-choice data with survey data, and document that 22% of students "self select," i.e., they do not rank their most-preferred school first. Out of these participants, 23% would have gotten into their favorite school if they had ranked it first. Under classical preferences, such ROLs are generically dominated and would require (wrong) knife-edge beliefs that attach probability zero to obtaining the skipped options, making the student indifferent between a truthful and self-selecting ROL. Such equilibria are not robust to minimal belief perturbations. Under report-dependent utility, self-selection can be rationalized as it entails a strategic trade-off akin to the immediate acceptance (Boston) mechanism. We capture self-selection by considering jump deviations that either move a less-preferred option forward or a more-preferred option backward in the ranking.

We contribute to the rich literature on strategy-proof mechanisms. The dominance of the truthful strategy for proposers in deferred-acceptance (DA) and top-trading cycles (TTC) mechanisms was established by Roth (1982a,b). As receivers might have an incentive to misrepresent preferences, two-sided strategic matching with incomplete information is complicated (Roth, 1989; Ehlers and Massó, 2007; Fernandez et al., 2021). We simplify the problem by focusing on the incentives of the proposing side, while inducing the receiving side to be truthful. The latter applies to settings in which this side is legally bound by objective priorities such as school choice. In a survey of the large experimental literature, Hakimov and Kübler (2021) document that truthfulness in DA and TTC is non-negligible and correlates with factors that do not impede strategy-proofness.

The economic literature mainly offers two strands of explanation. First, participants may fail to see the dominance of the truthful strategy and, hence, simply make a mistake in a complex mechanism. In this vein, there are efforts to make the strategy-proofness more apparent. For instance, Li (2017) introduces the concept of obvious strategy-proofness and, indeed, finds that truthfulness rates are higher in an obviously strategy-proof sequential serial dictatorship than in DA that does not have this property. However, the different performances of the two mechanisms can have alternative preference-based explanations such as our model. Somewhat at odds with explanations based on limited understanding is that misrepresentations persist in high-stakes environment with participants of high cognitive ability (Hassidim et al., 2017b; Rees-Jones and Skowronek, 2018; Shorrer and Sóvágó, 2017). A common misconception about strategy-proof mechanisms seems to be that participants falsely perceive a trade-off between preference and feasibility as in non-strategy-proof mechanisms. Katuščák and Kittsteiner (2020) find evidence in this direction in TTC and suggest an alternative framing nudging toward honesty. In our model, this trade-off originates in preferences rather than misunderstanding.

We contribute to a second branch of literature that rationalizes the"mistakes" as an optimal decision by a participant that fully understands the rules but has richer preferences. While Antler (2015) considers preferences that directly depend on the reported preferences of others, we consider preferences that directly depend on the own report. Dreyfuss et al. (2019) and Meisner and von Wangenheim (2021) study DA with expectation-based loss aversion, where proposers use the ROL to manage their expectations, which become their reference point. As in this paper, beliefs become crucial while they generically do not affect behavior in the classical framework. Similar to Proposition 2, Dreyfuss et al. (2019) construct parameters such that the true ROL is suboptimal with a reference-dependent utility function when the preference order does not coincide with the order of match probabilities. Although results appear similar, the desire to avoid disappointment with respect to expectations is a fundamentally different channel to drive misrepresentations. In such models, ROLs are evaluated only with respect to the lotteries over outcomes they generate, whereas in our model different ROLs that correspond to the same outcome lottery lead to different payoffs. Moreover, we investigate frustration from rejections or joy that reported preferences reciprocate which are independent of expectations.[3] Consequently, both theories can be disentangled. In contrast to Meisner and von Wangenheim (2021), who find that only top-choice monotone ROLs can be rationalized, we can for each ROL construct environments, in which this ROL is optimal.

2 The model

A participant in a direct strageyproof mechanism submits a rank-ordered list (ROL) that ranks n options from set $\mathcal{S}$. An ROL is a bijection $R\colon \mathcal{S} \to [\![1,n]\!] := \{1,\ldots,n\}$ that maps each option s into a rank $r \in [\![1,n]\!]$. Let $s_r^R = R^{-1}(r)$ be the r-th ranked option of some R, and we will sometimes display this function as a list, $R = (s_1^R, s_2^R, \ldots, s_n^R)$. If there is an outside option such as remaining unmatched, incomplete rankings are captured by ROLs that rank the "dropped" options after the outside option.

An entry of vector $\mathbf{v} = (v_s)_{s\in\mathcal{S}}$ represents the report-independent payoff from a match with option $s \in \mathcal{S}$. In addition, the participant receives report-dependent payoff $\rho(r)$ when she is assigned to her r-th ranked option, where ρ is a strictly decreasing function $\rho\colon [\![1,n]\!] \to \mathbb{R}$. Here, $\rho(1) > 0$ reflects the joy from experiencing no rejections and being accepted by the (reported) top choice, and $\rho(n) < 0$ reflects the chagrin from being rejected by every other option.

[3]Alternatively, ego-utility as formalized by Köszegi (2006) captures similar emotions, but there the self-regarding utility component inherently depends on beliefs about oneself.

Thus, the expected payoff from submitting ROL R is

$$U(\mathbf{v}|R) = \sum_{s \in \mathcal{S}} f^R_{R(s)} \big(v_s + \rho(R(s))\big) = \sum_{r=1}^{n} f^R_r \left(v_{s^R_r} + \rho(r)\right), \tag{1}$$

where f^R_r is the probability of matching with s^R_r under ROL R. To economize on notation, we will use accents to denote ROLs, and then let $s^{\widetilde{R}}_r = \widetilde{r}$, $f^{\widetilde{R}}_r = \widetilde{f}_r$, and $v_{s^{\widetilde{R}}_r} = v_{\widetilde{r}}$. Without loss of generality, we relabel $\mathcal{S} := \{1, \ldots, n\}$ with $v_1 \geq v_2 \geq \cdots \geq v_n = 0$,[4] and let the true ROL be denoted by $R = (1, 2, \ldots, n)$.

The beliefs about other participants' ROLs are taken as given. Given all others' ROLs (and priorities), let us call an option s attainable if our participant gets matched with s if she ranked it first. In any strategy-proof mechanism, she is matched with the highest-ranked attainable option.[5] While attainability probabilities are given by the other participants' ROLs, the options' priorities and capacities, and the specific mechanism's rules, the own ROL determines which attainable option is ranked highest and thereby generates the matching probabilities. Attainability probabilities are usually not independent even when all participants' preferences are independently drawn.

Let $A_s \in \{0, 1\}$ be a binary variable determining whether option s is attainable (1) or not (0). For any ranking $\widetilde{R}$, $\widetilde{f}_r = \Pr(A_{\widetilde{r}} = 1, A_{\widetilde{t}} = 0\, \forall t < r)$. If we interpret n as an outside option that is always attainable, i.e., it never rejects a participant and has unlimited capacity, no participant is ever assigned to an option ranked after n, $\widetilde{f}_r = 0$ for all $r > \widetilde{R}(n)$. Therefore, the order of options ranked $r > R(n)$ is irrelevant, and in this sense ranking an option after n corresponds to dropping it from the ranking.

Given $\mathbf{v}$, ρ and an attainability distribution P, we are interested in the optimal ROL R^* with

$$U(\mathbf{v}|R^*) \geq U(\mathbf{v}|\widetilde{R}) \quad \forall \widetilde{R} \neq R^*, \tag{2}$$

and we call R^* strictly optimal if all inequalities above are strict.

3 Analysis

Table 1 lists all possible attainability states and the corresponding payoff for each complete ROL with three options, $\mathcal{S} = \{1, 2, 3\}$. There is no safe outside option. Hence, all participants must rank all three options and will be matched to one of them. For each ROL, there are three states in which the participant ends up with her top choice, two states in which she is matched to her second choice, one state that matches her to her last choice. Which of these ROLs is optimal depends on the probability of each state, match utilities $\mathbf{v}$ and the report-dependent utility function ρ.

[4] This assumption implies that all options are preferred over the outside option n. This is innocuous because no participant would ever want to rank an unacceptable option before the

Attainability			1, 2, 3		1, 3, 2		2, 1, 3		2, 3, 1		3, 1, 2		3, 2, 1	
A_1	A_2	A_3	u_v	u_ρ	u_v	u_ρ	u_v	u_ρ	u_v	u_ρ	u_v	u_ρ	u_v	u_ρ
1	1	1	v_1	$\rho(1)$	v_1	$\rho(1)$	v_2	$\rho(1)$	v_2	$\rho(1)$	v_3	$\rho(1)$	v_3	$\rho(1)$
1	1	0	v_1	$\rho(1)$	v_1	$\rho(1)$	v_2	$\rho(1)$	v_2	$\rho(1)$	v_1	$\rho(2)$	v_2	$\rho(2)$
1	0	1	v_1	$\rho(1)$	v_1	$\rho(1)$	v_1	$\rho(2)$	v_3	$\rho(2)$	v_3	$\rho(1)$	v_3	$\rho(1)$
1	0	0	v_1	$\rho(1)$	v_1	$\rho(1)$	v_1	$\rho(2)$	v_1	$\rho(3)$	v_1	$\rho(2)$	v_1	$\rho(3)$
0	1	1	v_2	$\rho(2)$	v_3	$\rho(2)$	v_2	$\rho(1)$	v_2	$\rho(1)$	v_3	$\rho(1)$	v_3	$\rho(1)$
0	1	0	v_2	$\rho(2)$	v_2	$\rho(3)$	v_2	$\rho(1)$	v_2	$\rho(1)$	v_2	$\rho(3)$	v_2	$\rho(2)$
0	0	1	v_3	$\rho(3)$	v_3	$\rho(2)$	v_3	$\rho(3)$	v_3	$\rho(2)$	v_3	$\rho(1)$	v_3	$\rho(1)$

Table 1: All possible complete ROLs with three options and the corresponding payoffs in each possible attainability state. The report-independent utility u_v is listed on the left and the report-dependent utility u_ρ is listed on the right.

We start with the insight that for any ROL we can construct attainability distributions such that this ROL is optimal, and this is true for arbitrary report-dependent[6] and report-independent utilities.

Proposition 1. *For every ROL $\widetilde{R}$, there is an infinite set of attainability distributions $\widetilde{\mathcal{P}}$ such that for all $P \in \widetilde{\mathcal{P}}$, $\widetilde{R}$ is strictly optimal for every vector of report-independent utilities* $\mathbf{v}$ *and every strictly decreasing function ρ.*

The belief construction in the appendix is easy to illustrate with Table 1. Take an arbitrary ROL, say $\widetilde{R} = (2, 3, 1)$, and only consider the states in the fourth, sixth, and seventh line, i.e., states in which, aside from one, all options are unattainable. We see that state-by-state all complete ROLs garner the same payoff in the report-independent component. If we put all probability weight on state $(0, 1, 0)$, the participant is indifferent between ROLs $(2, 1, 3)$ and $(2, 3, 1)$ which she strictly prefers over all others. To make the weak preference strict, we now shift a sufficiently small probability mass p to state $(0, 0, 1)$. In this state, our ROL $(2, 3, 1)$ outperforms ROL $(2, 1, 3)$ such that it is strictly preferred in expectation. This p must not be too large as $p > \overline{p}$ could, for instance, render a deviation to ROL $(3, 2, 1)$ profitable in expectation. Calling $(0, \overline{p})$ an infinite set might be overly pompous, but it reflects that the construction is not a knife-edge case. Moreover, because the optimality is strict under the constructed beliefs, we can also sprinkle small probability masses ϵ over all other states while maintaining optimality. That is, the optimum is robust to small perturbations in the beliefs over all states. If there is a safe outside option, an additional constraint on p is necessary to prevent profitable deviations to truncated ROLs. We discuss in which settings such beliefs

safe outside option.

[5] Meisner and von Wangenheim (2021) introduce this terminology in their Lemma 1, which holds for all strategy-proof mechanisms.

[6] Our model has strictly decreasing functions ρ. If, for instance, $\rho(r) = \rho(n)$ for all $r \geq \overline{r}$, it is always weakly optimal to rank options in their true order from rank $\overline{r}$ onward. If $\overline{r} = 1$, we are in the standard setting without report-dependent utility such that the true ROL is weakly optimal for all beliefs.

can appear in the final section.

Proposition 1 should not be interpreted as an "anything-goes statement" voiding any predictive power of the model. While beliefs exist for each ROL to be optimal under arbitrary preferences, our theory predicts concrete ROLs to be optimal for given acceptability distributions and preferences. Next, we characterize beliefs such that the true ROL is always optimal. In words, submitting the true order implied by any given vector $\mathbf{v}$ is optimal for any ρ if and only if there does not exist any deviation that increases the probability of matching with the $\overline{r}$ top-ranked options for any $\overline{r}$.

Proposition 2. *Fix an arbitrary vector* $\mathbf{v}$ *and a non-truthful ROL* $\widetilde{R} = (\widetilde{1}, \ldots, \widetilde{n})$. *Then,* $U(\mathbf{v}|R) \geq U(\mathbf{v}|\widetilde{R})$ *for all decreasing functions* ρ *if and only if*

$$\sum_{r=1}^{\overline{r}} \big(P(A_r = 1, A_t = 0 \forall t < r) - P(A_{\widetilde{r}} = 1, A_{\widetilde{t}} = 0 \forall t < r) \big) \geq 0 \quad \forall \overline{r} \in [\![1, n]\!]. \quad (3)$$

Hence, the true ROL is optimal for every function ρ *if and only if the above inequalities hold against all non-truthful ROLs.*

Suppose condition (3) is violated for some $\widetilde{R}$ and some $\overline{r}$. If the participant would want to maximize the probability of being assigned to one of the top-ranked $\overline{r}$ options, she would prefer $\widetilde{R}$ over R. Intuitively, massively inflating $\rho(r)$ for all $r \leq \overline{r}$ leads to such incentives, and, consequently, ROL $\widetilde{R}$ yields a higher expected profit than the true ROL for some constructed functions ρ. If, to the contrary, all the inequalities of (3) hold, no change in ρ can upset this optimality of ordering options according to report-dependent utility.

Proposition 2 requires a high level of robustness for honesty in the sense that functional values of ρ can be arbitrarily large. This comes at a cost because every setting with a safe outside option violates (3) and thus allows that a truncated ROL can be optimal. For instance, a participant gets certain utility $\rho(1)$ from ranking the outside option first, while the utility of any other ROL is below $\widetilde{f}_1(v_1 + \rho(1)) + (1 - \widetilde{f}_1)(v_2 + \rho(2))$. As $\widetilde{f}_1 < 1$ for any non-outside option $\widetilde{1}$, a sufficiently high $\rho(1)$ makes ranking the safe option first optimal.

In general, comparing all possible ROLs can be tedious because attainability can be interdependent, implying the possibility of complicated profitable deviations. As we discuss in the next section, beliefs similar to the ones constructed for Proposition 1 suggest that low-priority participants rank options in order of attainability. However, also medium- and high-priority participants are documented submitting non-truthful ROLs. Among them, it is common to move forward just a few options in the ranking, often just one and to the top. We say $\widetilde{R}$ is an $\ell - k$-jump deviation from the true ROL R if the rank of some option $\ell > k$ is moved forward to $\widetilde{R}(\ell) = k$ and the options ranked afterwards in R move back by one, $\widetilde{R}(r) = r + 1$ for all $r \in [\![k, \ell - 1]\!]$. That is,

$$\widetilde{R} = (1, \ldots, k-1, \ell, k, \ldots, \ell - 1, \ell + 1, \ldots, m).$$

Only the colored ranks are affected as both ROLs list identical options at all ranks $r \notin [\![k, \ell]\!]$, i.e., $r = \widetilde{r}$ for all such r, while $\widetilde{k} = \ell$ and $\widetilde{r+1} = r$ for all $r \in [\![k, \ell-1]\!]$.

In all strategy-proof mechanisms, R and $\widetilde{R}$ generate identical match probabilities for each option ranked $r \notin [\![k, \ell]\!]$. Because the first $k-1$ proposals are identical, $f_r = \widetilde{f}_r$ for all $r < k$. The next $(\ell - k + 1)$ proposals differ but involve the same options in different order. At any step $t > \ell$, the participant is rejected by exactly the same options such that $f_r = \widetilde{f}_r$ for all $r > \ell$. Compared to $R, \widetilde{R}$ shifts more match probability weight to option $\ell = \widetilde{k}$ such that $\widetilde{f}_k = f_\ell + \delta_\ell^{R,\widetilde{R}}$ with $\delta_\ell^{R,\widetilde{R}} \geq 0$. This probability mass is shifted from the options which declined in the ranking such that for all $r \in [\![k, \ell-1]\!]$, we have $\widetilde{f}_{r+1} = f_r + \delta_r^{R,\widetilde{R}}$ with $\delta_r^{R,\widetilde{R}} \leq 0$ and $\sum_{r=k}^{\ell-1} \delta_r^{R,\widetilde{R}} = -\delta_\ell^{R,\widetilde{R}}$. Probability $\delta_r^{R,\widetilde{R}}$ is the probability that both r and ℓ are attainable, but no option ranked higher than r. The following lemma is true for any $\ell - k$-jump deviation from an arbitrary (not necessarily true) ROL.

Lemma 1. *The $\ell - k$-jump deviation $\widetilde{R}$ from ROL $\widehat{R}$ is strictly profitable, i.e., $U(\mathbf{v}|\widehat{R}) < U(\mathbf{v}|\widetilde{R})$, if and only if*

$$\sum_{r=k}^{\ell-1} \Big((\widehat{f}_r - \widehat{f}_\ell)(\rho(r) - \rho(r+1)) + \delta_r^{\widehat{R},\widetilde{R}} (\rho(k) - \rho(r+1)) \Big) < \sum_{r=k}^{\ell-1} \delta_r^{\widehat{R},\widetilde{R}} (v_{\widehat{r}} - v_{\widehat{\ell}}). \quad (4)$$

Inequality (4) reflects the trade-off between match utility and attainability probability. For example, it having an option ℓ with a high attainability probability "jump" over more preferred options that are most likely not attainable can be beneficial. In such a case, f_ℓ is large and f_r for the jumped options r are low. Moreover, the probability shifts $\delta_r \leq 0$ are also small. In combination, (4) holds, making the jump profitable. The inequality can also hold when $\delta_r = 0$ for all jumped options r, which distinguishes our theory from others where identical lotteries always yield the same utility. It can also be profitable to have an option ℓ with a high attainability probability jump options $r \in [\![k, \ell-1]\!]$ that are also likely attainable. The reason is that in such cases the probability shifts $|\delta_r|$ are large and the decrease on the left-hand side can be stronger than the increase on the right-hand side when preferences are not strong, i.e., when $(v_r - v_\ell)$ is small for all $r \in [\![k, \ell-1]\!]$.

Proposition 1 may be counter-intuitive. Since honesty is the best policy without report-dependent utility, one may expect to recover this property when this component approaches zero. This intuition can be maintained if the attainability distribution has full support in the sense that all attainability states have positive weight. In this case, sufficiently small report-dependent components imply that two adjacently ranked options must be in the order implied by $\mathbf{v}$. For any non-truthful ROL $\widehat{R}$, there must be some r such that $v_{\widehat{r+1}} > v_{\widehat{r}}$. Consider another ROL that swaps these two options. According to (4), this swap is profitable if

$$(\widehat{f}_r - \widehat{f}_{r+1} + \delta)(\rho(r) - \rho(r+1)) < \delta(v_{\widehat{r}} - v_{\widehat{r+1}}).$$

Since $\delta > 0$ under the full-support assumption, the right-hand side is negative and the left-hand side approaches zero as $(\rho(r) - \rho(r+1)) \to 0$. A series of such adjacent swaps culminates in the true ROL being optimal.

Corollary 1. *Suppose an attainability distribution P with a strictly positive weight on all attainability states. For all preferences $\mathbf{v}$, there exists a sufficiently small ϵ such that the true ROL is optimal if $(\rho(r) - \rho(r+1)) < \epsilon$ for all r.*

The proof follows from the argument above. However, reminiscent of Proposition 2, it is not true that sufficiently weak report-independent preferences imply a non-truthful ROL is optimal for all attainability distributions: even if $|v_s - v_{s'}| < \epsilon$ for all pairs $s, s' \in \mathcal{S}$, condition (3) guarantees that the truthful ROL is optimal for any $\epsilon > 0$.

4 Discussion

We have investigated the impact of report-dependent utility on behavior in strategy-proof mechanisms and established an inherent motive for self-selection. There are a plethora of sources for such a component of the payoff, such as self-regarding concerns, aversion to rejections, or signaling motives in a larger game. In our model, honesty can be guaranteed if and only if there is no conflict between where a participant wants to be assigned and what she finds feasible. We thus caution against taking reported preferences at face value for policy decisions, and we emphasize the importance of participants beliefs despite strategy-proofness. In the data, truthfulness is indeed negatively associated with the perceived attainability of preferred options. More research is necessary to identify whether this trade-off between preference and probability is preference-based or originates from misconceptions about the mechanism.

We show that preference misrepresentations can persist even as report-dependent utility becomes arbitrarily small. This result is only interesting if attainability distributions as constructed in Proposition 1 actually arise in reasonable settings. For example, consider a proposer in DA who knows that she has the lowest priority at every receiver, and suppose the capacities of all receivers sum up to the number of proposers. Since she is ranked last by all receivers, she will be matched to whoever the other proposers leave to her. For our proposer, there is only one attainable option in each possible state. Hence, it is optimal to rank options from most to least attainable.

Indeed, we see ROLs misrepresenting the induced report-independent preferences in this way in experimental studies of DA. For instance, a proposer in Li (2017, treatment SP-RSD) privately observes a priority score, an integer $i \in [\![1, 10]\!]$, and submits a complete ROL over four options with common values $\mathbf{v}$ for all participants. Indeed, only 61.1% of proposers with (the worst) priority score 1 submit the true ROL (1,2,3,4), and 17.8 % submit the dominated ROL (4,3,2,1) with the lowest possible payoff in the report-independent dimension. Since all proposers have the same $\mathbf{v}$, ranking options worst to best corresponds to the order of their likelihood of being left over.[7] The most common deviation among medium- and high-priority proposers is the simple jump deviation $(2, 1, 3, 4)$, which is in line with our arguments around Lemma 1.

Under such beliefs, the misrepresentations are not likely to have a large impact on allocations. However, in many settings, forming correct beliefs is complicated–even absent the usual biases in belief formation–because it is unclear how the other side evaluates the proposers. Such aggregate uncertainty is persistent and does not vanish as markets grow large. Our results put under scrutiny the alleged advantage that the success of strategy-proof mechanisms does not depend on beliefs. They also raise the question of whether such mechanisms really allocate the popular options to those participants who have the highest priorities or to those who merely think they do, when pessimistic high-priority participants shy away from applying. While we only considered a decision-theoretic view, it is easy to construct settings with multiple decision-makers such that preference misrepresentation persists in game-theoretic equilibrium. Moreover, such misrepresentation can be consequential in the sense that it impairs efficiency and stability with respect to report-independent utility. Self-selection is indeed consequential as documented by Chen and Pereyra (2019), where some self-selecting participants would have gotten into their favorite school, and Shorrer and Sóvágó (2017), where students leave scholarship money on the table.

At first glance, Proposition 2 seems to imply that honesty for all preference realizations cannot be obtained in any strategy-proof mechanism. It suggests that truthful ROLs can only be guaranteed for arbitrary report-dependent components if the individual preferences reverse the popular preferences, and this must be violated for most types by definition of popularity. Is there any setting in which report-dependent utility never causes issues with truthfulness in strategy-proof mechanisms? Yes, for instance, if all proposers in DA have preferences such that their v_s are individual iid draws and they all believe that receivers individually and privately draw priorities uniformly at random. Supposing that other proposers are truthful implies that each ROL is submitted with the same probability, which together with the receivers being ex-ante indifferent over all proposers implies that all options are equally likely to be attainable so that (3) holds with equality for all types. That is, in settings where preferences are maximally unknown such that a central mechanism collecting preferences has the largest benefit, report-dependent preferences do not cause problems in strategy-proof mechanisms. Pais and Pintér (2008) support our prediction as they find that truthfullness rates in DA and TTC are highest when participants know nothing about the others' preferences. In contrast to the standard model, our model can explain this observation: learning which options are likely to be contested can incentivize misrepresentations to avoid rejections from these options.

[7] Additionally, around 6.7% submit ROL (3,2,1,4), which can be explained similarly with an additional effect that in some settings $v_4 = \$0$ which participants may want to avoid. However, even in such settings some participants ranked option 4 first, which would be dominated even if they thought they played an immediate-acceptance mechanism.

If a market designer desires to implement allocations that are efficient or stable with respect to report-independent preferences, report-dependent payoffs may obstruct this goal. If misrepresentations are caused by disappointment aversion, it might be beneficial to tell participants that rejections are common in order to reduce the weight of gain-loss utility, parameter η in Dreyfuss et al. (2019) or Meisner and von Wangenheim (2021). Here, the effect of such an announcement is ambiguous. While $\rho(r)$ might increase for large r because rejections are perceived as less dramatic, $\rho(1)$ might also increase because a prevalence of rejections might lead to more pride in avoiding them. Alternatively, releasing information about the attainability of all options independent of the final allocation would make misrepresentations futile as a tool to avoid information about rejections. However, informing participants about rejections from options they did not even apply to may seem unnecessarily mean. In settings with a non-strategic market side with homogeneous preferences over participants, sequential serial dictatorship could reduce misrepresentations by letting participants choose sequentially in order of their priority as suggested by Li (2017) or Meisner and von Wangenheim (2021). When participants only select from a pool of options left once it is their turn to choose, the unattainability of preferred options does not influence their choice, and they can also credibly brag that they obtained their most-preferred option.

The fact that participants respond to advice appears to be incompatible with preference-based explanations. If the rules are fully understood, truthfulness rates should not increase when correct advice to report truthfully is provided, but they do. However, incorrect advice to self-select has an even larger effect in the opposite direction. For instance, the "wrong advice" in Guillen and Hing (2014)[8] is "Since the top schools will have many applicants you should be realistic and apply to schools where you are likely to gain acceptance. If your local school is quite good you should put it as your first preference." This advice is bad in terms of match utility, but it is good advice when participants care about how they ranked the school they end up with. The advice can be interpreted as a shift in mental focus from the report-independent to the report-dependent utility component.

Appendix

Proof of Proposition 1. Fix any arbitrary ROL $\widetilde{R} = (\widetilde{1}, \widetilde{2}, \dots, \widetilde{n})$, any function ρ, and any match utility vector $\mathbf{v}$. We construct an attainability distribution P such that $\widetilde{R}$ is strictly optimal. We assume that option n with $v_n = 0$ is a safe outside option, but the proof is straightforward to alter for the case without outside options.

The constructed P only puts positive weight on $\widetilde{R}(n)$ states. Let those weights and states be $q_{\widetilde{r}} = \Pr(A_{\widetilde{r}} = 1 = A_n, A_s = 0\, \forall s \neq \widetilde{r}, n)$, and let

$$q_{\widetilde{r}} > q_{\widetilde{r+1}} \quad \forall r \leq \widetilde{R}(n) \tag{5}$$

[8] They consider TTC. Similar observations exist for DA (Ding and Schotter, 2017, 2019).

with $\sum_{r=1}^{\widetilde{R}(n)} q_{\tilde{r}} = 1$. We first only compare $\widetilde{R}$ to ROLs $\widehat{R}$ of the same length as $\widetilde{R}$, i.e., $\widehat{R}(n) = \widetilde{R}(n)$, and note that

$$U(\mathbf{v}|\widetilde{R}) - U(\mathbf{v}|\widehat{R}) \geq \sum_{r=1}^{\widetilde{R}(n)} q_{\tilde{r}}\big(\rho(r) - \rho(\widehat{R}(\tilde{r}))\big), \tag{6}$$

because in each state both ROLs either yield the same report-independent utility $v_{\tilde{r}}$ or $\widehat{R}$ yields $v_n = 0 < v_{\tilde{r}}$ such that we can restrict attention to comparing report-dependent utility. Since ρ is decreasing and (5) holds, $\widetilde{R}$ puts the largest $\rho(r)$ on the most likely states. Hence, (6) is positive by the classical rearrangement inequality. Any longer ROL with $\widehat{R}(n) > \widetilde{R}(n)$ can only perform worse because it only additionally ranks options that are never attainable under P, which can only decrease report-dependent utility.

Next, we compare $\widetilde{R}$ to truncations of itself. Suppose $\widehat{R}(\tilde{r}) = \widetilde{R}(\tilde{r})$ for all $r < t < \widetilde{R}(n)$, and let $\widehat{R}(n) = t$. That is, $\widehat{R}$ lists the same options on ranks $r < t$ and drops all other options. Note that

$$U(\mathbf{v}|\widetilde{R}) \geq \sum_{r=1}^{t} q_{\tilde{r}}(v_{\tilde{r}} + \rho(r)) + \left(1 - \sum_{r=1}^{t} q_{\tilde{r}}\right) \rho(\widetilde{R}(n)) \quad \forall t < \widetilde{R}(n)$$

as $(v_{\tilde{r}} + \rho(r)) > (0 + \rho(\widetilde{R}(n))$ for all $r \in [\![t+1, \widetilde{R}(n)-1]\!]$. Hence, with $U(\mathbf{v}|\widehat{R}) = \sum_{r=1}^{t-1} q_{\tilde{r}}(v_{\tilde{r}} + \rho(r) - \rho(t)) + \rho(t)$, we have

$$U(\mathbf{v}|\widetilde{R}) - U(\mathbf{v}|\widehat{R}) \geq \left(1 - \sum_{r=1}^{t-1} q_{\tilde{r}}\right) \big(\rho(\widetilde{R}(n)) - \rho(t)\big) + q_{\tilde{t}}\big(v_{\tilde{t}} + \rho(t) - \rho(\widetilde{R}(n))\big),$$

which is positive for all t if

$$q_{\tilde{t}} \geq \left(1 - \sum_{r=1}^{t-1} q_{\tilde{r}}\right) \frac{\rho(t) - \rho(\widetilde{R}(n))}{v_{\tilde{t}} + \rho(t) - \rho(\widetilde{R}(n))} = \left(1 - \sum_{r=1}^{t-1} q_{\tilde{r}}\right) \alpha \quad \forall t < \widetilde{R}(n), \tag{7}$$

where $\alpha \in (0, 1)$ because $\rho(t) > \rho(\widetilde{R}(n))$ for all $t < \widetilde{R}(n)$. If additionally (5) holds, also all other truncated ROLs of length t yield a lower expected payoff than $\widetilde{R}$.

Let set $\widetilde{\mathcal{P}}$ be the set of all P constructed as above, such that both (5) and (7) hold, which is non-empty and infinite. $\square$

Proof of Proposition 2. First, note that

$$U(\mathbf{v}|R) - U(\mathbf{v}|\widetilde{R}) = \sum_{r=1}^{n} \big(v_r(f_r - \widetilde{f}_{\widetilde{R}(r)}) + \rho(r)(f_r - \widetilde{f}_r)\big) = \Delta_v + \Delta_\rho,$$

where $\Delta_v > 0$ as strategy-proofness implies a first-order stochastic dominance of the true lottery with respect to the report-independent utility.

Suppose (3) is violated for some $\overline{r}$ of ROL $\widetilde{R}$, and let the difference in (3) be $\Delta_{\overline{r}} < 0$. We construct a decreasing ρ such that $\rho(r) \to \rho(1)$ for all $r \leq \overline{r}$ and $\rho(r) \to 0$ for all $r > \overline{r}$. Then, we have

$$U(\mathbf{v}|R) - U(\mathbf{v}|\widetilde{R}) \to \Delta_v + \rho(1) \sum_{r=1}^{\overline{r}} (f_r - \widetilde{f}_r) + 0 = \Delta_v + \rho(1)\Delta_{\overline{r}},$$

which can be made arbitrarily negative by increasing $\rho(1) > {}^{-\Delta_v}/_{\Delta_{\overline{r}}} > 0$. Hence, there are functions ρ such that $U(\mathbf{v}|R) < U(\mathbf{v}|\widetilde{R})$.

Suppose (3) holds for all $\overline{r}$, and fix any arbitrary $\mathbf{v}$ and ρ. Under strategy-proofness, $U(\mathbf{v}|R) - U(\mathbf{v}|\widetilde{R}) \geq \Delta_\rho$, and we see that (3) implies

$$\begin{aligned}
\Delta_\rho &= \sum_{r=1}^{n-1} (f_r - \widetilde{f}_r)\rho(r) + \rho(n) \left(\left(1 - \sum_{r=1}^{n-1} f_r\right) - \left(1 - \sum_{r=1}^{n-1} \widetilde{f}_r\right)\right) \\
&= \sum_{r=1}^{n-1} (f_r - \widetilde{f}_r)(\rho(r) - \rho(n)) \\
&= \sum_{r=1}^{n-1} (f_r - \widetilde{f}_r) \sum_{i=r}^{n-1} (\rho(i) - \rho(i+1)) \\
&= \sum_{r=1}^{n-1} (\rho(r) - \rho(r+1)) \left(\sum_{i=1}^{r} f_i - \sum_{i=1}^{r} \widetilde{f}_r \right) > 0,
\end{aligned}$$

as for each r the first factor is positive for any decreasing ρ and the second factor is positive when (3) holds. $\square$

Proof of Lemma 1. By definition, $U(\mathbf{v}|\widehat{R}) - U(\mathbf{v}|\widetilde{R}) < 0$ if and only if

$$\sum_{r=1}^{m} \left(\widehat{f}_r \left(v_{\widehat{r}} + \rho(r)\right) - \widetilde{f}_r \left(v_{\widetilde{r}} + \rho(r)\right)\right) < 0$$

$$\sum_{r=k}^{\ell} \left(\widehat{f}_r \left(v_{\widehat{r}} + \rho(r)\right) - \widetilde{f}_r \left(v_{\widetilde{r}} + \rho(r)\right)\right) < 0$$

$$\sum_{r=k}^{\ell-1} \left(\widehat{f}_r \left(v_{\widehat{r}} + \rho(r)\right) - (\widehat{f}_r + \delta_r^{\widehat{R},\widetilde{R}}) \left(v_{\widehat{r}} + \rho(r+1)\right)\right)$$
$$+ \widehat{f}_\ell(v_{\widehat{\ell}} + \rho(\ell)) - (\widehat{f}_\ell + \delta_\ell^{\widehat{R},\widetilde{R}}) \left(v_{\widehat{\ell}} + \rho(k)\right) < 0$$

$$\sum_{r=k}^{\ell-1} \left(\widehat{f}_r \left(\rho(r) - \rho(r+1)\right) - \delta_r^{\widehat{R},\widetilde{R}} \left(v_{\widehat{r}} + \rho(r+1)\right)\right)$$
$$+ \widehat{f}_\ell(\rho(\ell) - \rho(k)) - \delta_\ell^{\widehat{R},\widetilde{R}} \left(v_{\widehat{\ell}} + \rho(k)\right) < 0$$

$$\sum_{r=k}^{\ell-1} \left(\widehat{f}_r(\rho(r) - \rho(r+1)) - \delta_r^{\widehat{R},\widetilde{R}} \rho(r+1)\right) + \widehat{f}_\ell(\rho(\ell) - \rho(k)) - \delta_\ell^{\widehat{R},\widetilde{R}} \rho(k) < \sum_{r=k}^{\ell} \delta_r^{\widehat{R},\widetilde{R}} v_{\widehat{r}}.$$

Because $(\rho(k) - \rho(\ell)) = \sum_{r=k}^{\ell-1}(\rho(r) - \rho(r+1))$ and $-\delta_{\ell}^{\widehat{R},\widetilde{R}} = \sum_{r=k}^{\ell-1} \delta_{r}^{R,\widetilde{R}}$, we can rewrite the above as (4). $\square$

References

Antler, Y., 2015. Two-sided matching with endogenous preferences. American Economic Journal: Microeconomics 7 (3), 241–58.

Börgers, T., 2015. An Introduction to the Theory of Mechanism Design. Oxford University Press, USA.

Chen, L., Pereyra, J. S., 2019. Self-selection in school choice. Games and Economic Behavior 117, 59–81.

Chen, Y., Sönmez, T., 2006. School choice: An experimental study. Journal of Economic Theory 127 (1), 202–231.

Ding, T., Schotter, A., 2017. Matching and chatting: An experimental study of the impact of network communication on school-matching mechanisms. Games and Economic Behavior 103, 94–115.

Ding, T., Schotter, A., 2019. Learning and mechanism design: An experimental test of school matching mechanisms with intergenerational advice. The Economic Journal 129 (623), 2779–2804.

Dreyfuss, B., Heffetz, O., Rabin, M., 2019. Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. American Economic Journal: Microeconomics.

Ehlers, L., Massó, J., 2007. Incomplete information and singleton cores in matching markets. Journal of Economic Theory 136 (1), 587–600.

Fernandez, M., Rudov, K., Yariv, L., 2021. Centralized matching with incomplete information.

Guillen, P., Hing, A., 2014. Lying through their teeth: Third party advice and truth telling in a strategy proof mechanism. European Economic Review 70, 178–185.

Hakimov, R., Kübler, D., 2021. Experiments on centralized school choice and college admissions: A survey. Experimental Economics 24 (2), 434–488.

Hassidim, A., Marciano, D., Romm, A., Shorrer, R. I., 2017a. The mechanism is truthful, why aren't you? American Economic Review 107 (5), 220–24.

Hassidim, A., Romm, A., Shorrer, R. I., 2017b. Redesigning the israeli psychology master's match. American Economic Review 107 (5), 205–09.

Katuščák, P., Kittsteiner, T., 2020. Strategy-proofness made simpler. Tech. rep., working paper.

Köszegi, B., 2006. Ego utility, overconfidence, and task choice. Journal of the European Economic Association 4 (4), 673–707.

Li, S., 2017. Obviously strategy-proof mechanisms. American Economic Review 107 (11), 3257–87.

Meisner, V., von Wangenheim, J., 2021. School choice and loss aversion. Tech. rep., working paper.

Pais, J., Pintér, Á., 2008. School choice and information: An experimental study on matching mechanisms. Games and Economic Behavior 64 (1), 303–328.

Rees-Jones, A., Skowronek, S., 2018. An experimental investigation of preference misrepresentation in the residency match. Proceedings of the National Academy of Sciences 115 (45), 11471–11476.

Roth, A. E., 1982a. The economics of matching: Stability and incentives. Mathematics of Operations Research 7 (4), 617–628.

Roth, A. E., 1982b. Incentive compatibility in a market with indivisible goods. Economics Letters 9 (2), 127–132.

Roth, A. E., 1989. Two-sided matching with incomplete information about others' preferences. Games and Economic Behavior 1 (2), 191–209.

Shorrer, R. I., Sóvágó, S., 2017. Obvious mistakes in a strategically simple college admissions environment.