Report-dependent utility and strategy-proofness*

Vincent Meisner†

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 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. We substantiate the theory with already existing evidence and provide novel testable predictions.

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

*I thank Inácio Bó, Rustamdjan Hakimov, Timm Opitz, Jinju Rhee, Christoph Schwaiger, Jonas von Wangenheim, Georg Weizsäcker. and seminar participants in Berlin, Belfast, and Padova for useful comments. 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. 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 a matching mechanism that is strategy-proof with respect to standard preferences. 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[2] or because it is imposed by the other market side.[3]

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 participant might also be interested in signaling to the receivers that her preferences are in line with theirs.

[3] For example, some private universities in the centralized admission program in Turkey offer "preference scholarships," which reduce tuition for students that rank them among the top ranks.

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 the 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 assign probability zero to obtaining the skipped options, making the student indifferent between a truthful and a self-selecting ROL. Such equilibria are not robust to minimal belief perturbations. Under report-dependent utility, self-selection can be rationalized as such preferences entail a strategic trade-off akin to the immediate acceptance (Boston) mechanism. We capture self-selection by considering jump and swap 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). If only one side of the mechanism is strategy-proof, two-sided strategic matching with incomplete information is complicated (Roth, 1989; Ehlers and Massó, 2007; Fernandez et al., 2022). We focus on the incentives of the strategy-proof side, while inducing the other side to be truthful, e.g., through objective priorities such as in 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. In Section 3.1, we discuss these observations in the light of report-dependent preferences. We predict (i) a pattern of misrepresentations under common values consistent with Li (2017), (ii) that weak preferences trigger manipulations consistent with Klijn et al. (2013), and (iii) how more information about priorities can reduce truthfulness rates consistent with Pais and Pintér (2008).

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 its static version 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 environments with participants of high cognitive ability (Hassidim et al., 2017b; Rees-Jones and Skowronek, 2018; Shorrer and Sóvágó, 2017). Katuščák and Kittsteiner (2020) find evidence that participants in TTC falsely perceive a trade-off between preference and feasibility, and they propose an alternative framing that nudges 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. Although results appear similar, the desire to avoid disappointment with respect to expectations is a fundamentally different channel to drive misrepresentations. The frustration from rejections or the joy that reported preferences reciprocate is independent of expectations.[4] Our predictions in Section 3.1 enable us to differentiate between both theories.

2 The model

A participant in a matching mechanism submits a rank-ordered list (ROL) that ranks n options from set $\mathcal{S}$. This mechanism is strategy-proof with respect to standard (report-independent) preferences, and we always use this term referring to standard preferences. An ROL is a bijection $R\colon \mathcal{S} \to [\![1,n]\!] := \{1,\dots,n\}$ 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, $R = (s_1^R, s_2^R, \dots, s_n^R)$.

An entry of vector $\mathbf{v} = (v_s)_{s\in\mathcal{S}} \in \mathbb{R}^n$ 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.

Thus, the expected payoff from submitting ROL R is

$$U_\rho(\mathbf{v}|R) = \sum_{s\in\mathcal{S}} f_{R(s)}^R\big(v_s + \rho(R(s))\big) = \sum_{r=1}^{n} f_r^R\left(v_{s_r^R} + \rho(r)\right), \tag{1}$$

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 $s_r^{\widetilde{R}} = \widetilde{r}$, $f_r^{\widetilde{R}} = \widetilde{f}_r$, and

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

$v_{s_{\widetilde{r}}^{\widetilde{R}}} = v_{\widetilde{r}}$. Without loss of generality, we relabel $\mathcal{S} := \{1, \ldots, n\}$ with $v_1 \geq v_2 \geq \cdots \geq v_n$, and we let the "true" ROL be denoted by $R = (1, 2, \ldots, n)$.

For a fixed mechanism and given all others' ROLs, priorities and capacities, we call an option s attainable if there exists some ROL R such that the given mechanism assigns our participant to option s. In a strategy-proof mechanism, this is the case if and only if she is assigned to s when ranking it first. Let $A_s \in \{0, 1\}$ be a binary variable determining whether option s is attainable (1) or not (0). The attainability distribution P is a probability distribution over attainability states $(A_s)_{s \in \mathcal{S}}$. Since attainability is a key concept in this paper, we open this black box in an exemplary setting in the appendix for three well-known strategy-proof mechanisms: deferred-acceptance (DA), top-trading cycles (TTC), and serial dictatorship (SD). By Lemma 2 in the appendix, a strategy-proof mechanism matches our participant to her highest-ranked attainable option. For any ranking $\widetilde{R}$,

$$\widetilde{f}_r = \Pr(A_{\widetilde{r}} = 1, A_{\widetilde{t}} = 0 \, \forall t < r). \tag{2}$$

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

$$U_\rho(\mathbf{v}|R^*) \geq U_\rho(\mathbf{v}|\widetilde{R}) \quad \forall \widetilde{R} \neq R^*, \tag{3}$$

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

Working with the reduced attainability framework also permits the definition of an outside option and thereby allows for truncated ROLs. An outside option $o \in \mathcal{S}$ can be a fictional option that never rejects the participant such as remaining unmatched (also called getting matched to oneself). In our reduced from, this means that o is always attainable, regardless of the other participants' ROLs. It does not matter if o is a (fictional) option that has unlimited capacity or, for example, a district school at which the participant is a student with the highest priority. An outside option in this sense is also dependent on the mechanism. To illustrate, for the first chooser in SD, all schools are essentially outside options. Because no participant is ever assigned to an option ranked behind such an option o, $\widetilde{f}_r = 0$ for all $r > \widetilde{R}(o)$. Therefore, the order of options ranked $r > R(o)$ is irrelevant, and in this sense ranking an option worse than o corresponds to dropping it from the ranking. In the following, we only consider ROLs listing only acceptable options that are preferred over a possible outside option o. Any ROL listing an unacceptable option s with $v_s < v_o$, i.e., any ROL $\widetilde{R}$ with $\widetilde{R}(s) < \widetilde{R}(o)$, can be improved upon by dropping s behind o in the ranking.

3 Analysis

In this section, we first characterize which ROLs can be rationalized under report-dependent preferences, and we characterize for which attainability distributions the truthful ROL is optimal. Finally, we use our insights to formulate predictions based on the model. To illustrate our first result, consider Table 1, which 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. For each ROL,

there are four 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, report-independent utilities $\mathbf{v}$, and the report-dependent utility function ρ.

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 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 strictly decreasing[5] report-dependent and report-independent utilities. If an outside option exists, we fix it and only alter the attainability distribution regarding options that are acceptable with respect to this outside option.

Proposition 1. *For every ROL $\widetilde{R}$ listing only acceptable options, there is an attainability distribution $\widetilde{P}$ such that $\widetilde{R}$ is strictly optimal for every vector of report-independent utilities $\mathbf{v}$ and every function ρ given any attainability distribution P in an open ball around $\widetilde{P}$.*

The construction of $\widetilde{P}$ 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. Here, we see that state-by-state all 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 over $(2, 1, 3)$ 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. Since optimality is strict given such a $\widetilde{P}$ and expected utility is continuous in P, we can construct an open environment around $\widetilde{P}$ while maintaining optimality, which reflects that the construction is not a knife-edge case. Because we can also sprinkle small probability masses ϵ over all other

[5]If, for instance, $\rho(r) = \rho(n)$ for all $r \geq \overline{r}$, it is always weakly optimal to rank options in the 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 dominant.

states while maintaining optimality, the optimum is robust to small perturbations in the distribution over all states. If there is a safe outside option, an additional constraint on p is necessary to prevent profitable deviations to truncated ROLs. In Prediction 1 in Section 3.1, we again pick up the intuition behind this result.

According to Proposition 1, we can construct attainability distributions such that any ROL of acceptable options becomes optimal regardless of how the true (i.e., report-independent) preferences rank these options. While we can modify beliefs on priorities and ROLs of the other players such that a participant in DA or TTC faces the attainability distribution constructed above, this is not true for all strategy-proof mechanisms. For instance, a lottery that ignores all priorities and ROLs and allocates options at random is strategy-proof, but by construction beliefs about ROLs or priorities have no impact on the fixed attainability distribution as seen in the example in the appendix. Alternatively, consider the first chooser in SD.[6] Because this participant always gets what she ranked first independently of others' ROLs, all options are always attainable, making the true top-choice an endogenous outside option such that all others are unacceptable. Altering her beliefs about other player's ROLs has no impact on the attainability distribution such that this participant will rank her true top-choice first for all beliefs.

Moreover, some information environments impose restrictions on attainability. For example, designating a priority option (such as a district school) essentially means imposing that it is always attainable. That is, it is an endogenous outside option and therefore determines which options are acceptable. By Proposition 1, we can rationalize moving the outside option upwards in the ranking, but we cannot rationalize moving it downwards as this would imply listing an unacceptable option, which is never optimal. To illustrate, consider a proposing student in DA who knows to have the highest priority at their district school and the lowest priority everywhere else. For this student, we can construct beliefs about the submissions of other students such that any ROL that only ranks schools weakly preferred over the district school is optimal. However, we cannot construct beliefs such that this student would like to rank an unacceptable option, i.e., a school they consider strictly worse that their district school. Similarly, conditional on the information that some school is the district school, we cannot construct an attainability distribution such that this school is not attainable: any belief consistent with such an attainability distribution would contradict the condition that our student has highest priority at the district school.

In the next proposition, we characterize attainability distributions such that the true ROL is always optimal. Submitting the true order implied by any given vector $\mathbf{v}$ is optimal for any ρ if and only if (4) holds. In words, this condition means that there does not exist any deviation that increases the probability of matching with the $\overline{r}$ top-ranked options for any $\overline{r}$. It describes the attainability distributions such that $\sum_{r=1}^{\overline{r}} f_r \geq \sum_{r=1}^{\overline{r}} \widetilde{f}_r$ for any ROL $\widetilde{R}$ and all $\overline{r} \in [\![1, n]\!]$.

[6] In our decision-theoretic setting, being first chooser in SD is another mechanism than being last chooser or having a random order in SD.

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

$$\sum_{r=1}^{\overline{r}} \big(P(A_r = 1, A_t = 0 \forall t < r) - P(A_{\widetilde{r}} = 1, A_{\widetilde{t}} = 0 \forall t < r)\big) \geq 0 \quad \forall \overline{r} \in [\![1, n]\!]. \tag{4}$$

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

Suppose condition (4) is violated for some $\widetilde{R}$ and $\overline{r} = 1$. This means that the participant's most-preferred option is not the most attainable option. There is another option such that ranking it first yields a higher probability of assignment to the (reported) top choice than when ranking the true favorite first. That is, there exists some $\widetilde{R}$ such that $\widetilde{f}_1 > f_1$. For example, this is true if there is a safe outside option $o \neq 1$. In this case, a participant gets certain utility $\rho(1) + v_o$ from ranking option $o = \widetilde{1}$ first, while the expected utility from the true ROL is below $f_1(v_1 + \rho(1)) + (1 - f_1)(v_2 + \rho(2))$. As $f_1 < \widetilde{f}_1 = 1$ for any non-outside option favorite, we can set a sufficiently high $\rho(1)$ (and low $\rho(2)$) to make ranking the safe option first optimal. That is, Proposition 2 requires a high level of robustness for honesty in the sense that functional values of ρ can be arbitrarily large.

Similar constructions of ρ can make a non-truthful deviation $\widetilde{R}$ profitable whenever (4) is violated for any $\overline{r} > 1$. Intuitively, massively inflating $\rho(r)$ for all $r \leq \overline{r}$ leads to incentives such that maximizing the probability of being assigned to one of the $\overline{r}$ highest-ranked options becomes of first-order importance. Consequently, ROL $\widetilde{R}$ yields a higher expected profit than the true ROL for some constructed functions ρ for any violation of (4). If, to the contrary, all the inequalities of (4) hold, no decreasing function ρ can upset the optimality of ordering options according to the given $\mathbf{v}$.

In general, comparing all possible ROLs can be tedious because attainability can be interdependent, implying the possibility of complicated profitable deviations. We now focus on popular deviations commonly observed in the data. We capture self-selection strategies with jump deviations that simply move forward one option in the ranking. In fact, according to Hakimov and Kübler (2021, Section 3.4.1), a special case of a jump deviation, simply swapping the first two options in the true ROL, is the modal manipulation in many studies. We say $\widetilde{R}$ is an ℓ-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 worse in R move down by one rank, $\widetilde{R}(r) = r + 1$ for all $r \in [\![k, \ell - 1]\!]$. That is,

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

Only underlined 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 $r \notin [\![k, \ell]\!]$. For example in DA, the first $k-1$ proposals are identical, implying $f_r = \widetilde{f}_r$ 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 under both ROLs 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 mass $\delta_r^{R,\widetilde{R}}$ is the probability that both r and ℓ are attainable, while each option ranked better than r is unattainable. The following lemma is true for any ℓ-k-jump deviation from an arbitrary (not necessarily true) ROL.

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

$$\sum_{r=k}^{\ell-1} \Big((\widehat{f}_r - \widehat{f}_\ell)(\rho(r) - \rho(r+1)) + \delta_r^{\widehat{R},\widetilde{R}} (\rho(k) - \rho(r+1)) \Big) < \sum_{r=k}^{\ell-1} \delta_r^{\widehat{R},\widetilde{R}} (v_{\widehat{r}} - v_{\widehat{\ell}}). \quad (5)$$

Inequality (5) is an algebraic rearrangement of $U_\rho(\mathbf{v}|\widehat{R}) < U_\rho(\mathbf{v}|\widetilde{R})$, and it reflects the trade-off between match utility and attainability probability: In a profitable jump deviation, the loss (or gain) in report-independent payoff on the right-hand side is compensated by the gain (or loss) in the report-dependent payoff on the left-hand side. For example, having an option ℓ with a high attainability probability "jump" over more preferred options that are almost unattainable can be beneficial. In such a case, f_ℓ is large and $f_r \approx 0$ for the jumped options r. Moreover, the probability shifts $\delta_r \approx 0$ are also small. In combination, (5) holds, making the jump profitable. 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 as function ρ becomes close to constant. This intuition can be maintained if the attainability distribution has full support in the sense that all attainability states have positive weight. For any non-truthful ROL $\widehat{R}$, there must be some pair of options that is adjacently ranked in the order reversing $\mathbf{v}$, i.e., there is some ℓ such that $v_{\widehat{\ell-1}} < v_{\widehat{\ell}}$. Consider another ROL that swaps these two options such that they do reflect the order of $\mathbf{v}$. According to (5), this $(\ell - 1)$-ℓ-swap is profitable if

$$(\widehat{f}_{\ell-1} - \widehat{f}_\ell + \delta_{\ell-1}^{\widehat{R},\widetilde{R}})(\rho(\ell - 1) - \rho(\ell)) < \delta_{\ell-1}^{\widehat{R},\widetilde{R}} (v_{\widehat{\ell-1}} - v_{\widehat{\ell}}) = \delta_\ell^{\widehat{R},\widetilde{R}} (v_{\widehat{\ell}} - v_{\widehat{\ell-1}}). \quad (6)$$

Since $-\delta_{\ell-1}^{\widehat{R},\widetilde{R}} = \delta_\ell^{\widehat{R},\widetilde{R}} > 0$ under the full-support assumption and $v_{\widehat{\ell}} > v_{\widehat{\ell-1}}$, the right-hand side is strictly positive, while the left-hand side approaches zero as

$(\rho(\ell-1)-\rho(\ell)) \to 0$. A series of such adjacent swaps culminates in the true ROL being optimal.

Corollary 1. *Consider an attainability distribution P with a strictly positive weight on all attainability states. For all $\mathbf{v}$, there exists a sufficiently small $\epsilon > 0$ 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 $\mathbf{v}$ imply a non-truthful ROL is optimal: even if $|v_s - v_{s'}| < \epsilon$ for all pairs $s, s' \in \mathcal{S}$, condition (4) guarantees that the truthful ROL is optimal for any $\epsilon > 0$. Reminiscent of Proposition 1, Corollary 1 needs the full-support condition.

We continue to consider $(\ell-1)$-ℓ-swaps as above, since 1-2-swaps in the true ROL are the most popular manipulation. Inequality (6) tells us when an $(\ell-1)$-ℓ-swap in the true ROL R is unprofitable. In such a swap from R to $\widehat{R}$, $\widehat{\ell-1} = \ell$ and $\widehat{\ell} = \ell - 1$, we have $f_{\ell-1} = \widehat{f}_\ell - \delta_{\ell-1}^{\widehat{R},R} = \widehat{f}_\ell + \delta_\ell^{\widehat{R},R}$ so that (6) can be expressed as

$$\frac{\widehat{f}_{\ell-1} - f_{\ell-1}}{\delta_\ell^{\widehat{R},R}} < \frac{v_{\ell-1} - v_\ell}{\rho(\ell-1) - \rho(\ell)}. \tag{7}$$

Since the right-hand side and $\delta_\ell^{\widehat{R},R}$ are positive, this inequality always holds when $\widehat{f}_{\ell-1} < f_{\ell-1}$. In words, swapping two options $\ell-1$ and ℓ in the true ROL is always unprofitable when it decreases the probability of matching with the option on rank $\ell - 1$.

Corollary 2. *If the optimal ROL is an $(\ell-1)$-ℓ-swap $\widehat{R}$ of the true ROL R, it must be that this deviation increases the probability of assignment to the $(\ell-1)$-th ranked option, $\widehat{f}_{\ell-1} > f_{\ell-1}$. Otherwise, ${}^{(v_{\ell-1}-v_\ell)}/_{(\rho(\ell-1)-\rho(\ell))}$ must be sufficiently large.*

Such a deviation can be in line with the priority-option bias or the small-option bias discussed in the next section. However, the insight above is more general. If, compared to option $\ell-1$, ℓ is (perceived to be) less popular among competitors with higher priority, swapping those options in the ranking can be profitable even when option ℓ is smaller that option $\ell-1$ or when the participant has low priority at both options.

3.1 Predictions

Our results put under scrutiny the alleged advantage that the success of strategy-proof mechanisms does not depend on beliefs. However, they are only interesting if the constructed attainability distributions actually arise from reasonable beliefs in mechanisms in use. Proposition 1 should not be interpreted as an "anything-goes statement" voiding any predictive power of the model. While attainability distributions exist for each ROL to be optimal under any $\mathbf{v}$ and ρ, our theory predicts concrete ROLs to be optimal for given attainability distributions and

preferences. In this section, we provide some testable predictions of our model, and we discuss experimental evidence consistent with the predictions in the online appendix.

In the proof of Proposition 1, we fix an ROL and then construct an attainability distribution such that this ROL is strictly optimal. This construction is easiest to illustrate in SD without an outside option when there are in total n participants, capacities sum up to n, and our participant is last to choose. If she knew the reported ROLs of all others, she would optimally rank first the option that is left over—let us call it a_1—as she is assigned to this option with certainty and for all her ROLs. Similarly, if there is a small probability that instead another option a_2 will be left over, it would optimally be ranked second, and so on. That is, if a_k is the k-th likely option to be left over, it is the k-th most attainable option, and the optimal ROL of the final chooser in SD is $(a_1, a_2, \ldots, a_n)$, independent of her preferences. If we consider a participant in DA or TTC who knows to have the lowest priority, an analogous logic applies. In DA, such a participant gets what others do not want because she gets rejected whenever another proposer approaches her tentative match. In TTC, such a participant gets what others do not want because no option points at her as long as other participants are present. Hence, in all three mechanisms and for every combination of the other participants' ROLs, the state is such that only one option is attainable. Essentially, this is the construction of the attainability distribution in Proposition 1, and it implies the following prediction.

Prediction 1. *Suppose all options have unit capacity and there are n options and participants. All n participants have a common preference vector $\mathbf{v}$.*[7] *Consider a participant who knows to have the lowest priority at all options. In DA, TTC, or (priority-ordered) SD, and for any $\mathbf{v}$ and any ρ, this participant optimally ranks options from most to least attainable.*

Indeed, Li (2017, treatment SP-RSD) records non-truthful deviations that are consistent with Prediction 1. Unfortunately, the popularity of ROL $(4, 3, 2, 1)$ in his common-value setting confounds two preference-based explanations. This ROL is also the only top-choice monotone[8] ROL that starts with the most attainable option. Hence, in keeping with Meisner and von Wangenheim (2021, Proposition 1), this ROL is also rationalizable for the lowest-priority agent under expectation-based loss aversion (EBLA). In the online appendix, we discuss how the experiment can be modified to disentangle the theories.

Li's common value setting with common priority rankings gives rise to additional predictions. As in many other studies, the most common manipulation in Li (2017) is the 1-2-swap $(2, 1, 3, 4)$. If we consider the first-ranked options in the

[7] We allow for indifference, $v_k = v_{k+1}$ and, hence, we can split up an option s with capacity $q_s > 1$ into q_s separate options with unit capacity over which the participants are indifferent.

[8] Meisner and von Wangenheim (2021, Proposition 1) show that only such ROLs can be optimal in their setting. An ROL is top-choice monotone if it reverses the order of options preferred to the reported top choice and preserves the order of the other options.

modal deviations for each priority score separately, we observe a monotonicity: lower scores tend to rank worse options first. A line of reasoning behind such deviations becomes clear in Prediction 2, which slightly alters the informational setting of Li (2017).

Prediction 2. *Suppose all options have unit capacity and there are n options and participants. All n participants have a common preference vector $\mathbf{v}$. All options have the same priority ranking over participants, and each participant k knows to have the k-th priority. In DA, TTC, or SD (in order of priority) and for any ρ, this participant optimally ranks the k-th preferred option (according to $\mathbf{v}$) first.*

Here, all options are attainable to the highest-priority participant 1 who just chooses her favorite, and all ROLs ranking option 1 first are payoff-equivalent. Given the behavior of the higher-priority participants, participant k essentially selects her final match with certainty. In this Nash equilibrium, also participants with standard preferences or EBLA optimally submit such a report. However, in both these cases the participant is indifferent between an ROL with the predicted structure and the true ROL or any other ROL only ranking options $s < k$ already selected by others. In contrast, under report-dependent utility the preference is strict because the match is certain and $\rho(1) > \rho(r)$ for all $r > 1$. Consequently, a participant is willing to pay up to $(\rho(1)-\rho(k)) > 0$ to perform a jump manipulation in the true ROL. In general, we can distinguish our theory from others that simply treat ROLs as lotteries over match outcomes such that identical lotteries yield the same utility. Opposed to such approaches, a jump deviation can be profitable even when $\delta_r = 0$ for all jumped options r, see (5).

According to our model, participants in Li (2017) apply the logic behind Prediction 2 probabilistically as they can only imperfectly infer their priority rank from their privately observed priority score. Hence, participants with medium or high priority scores and either sufficiently large $(\rho(1) - \rho(2))$ (or sufficiently small $(v_1 - v_2)$) prefer ROL $(2, 1, 3, 4)$ over the true ROL, which brings us to our next prediction. Based on (7) with $v_1 \approx v_2$, this prediction gives conditions conducive to swaps that are consistent with the priority-option bias, the small-option bias, and the similar-preference bias coined by Chen and Sönmez (2006).

Prediction 3. *Consider a participant and her two most preferred options 1 and 2, and let the report-independent preferences over the two be weak, i.e., $v_1 = v_2 + \epsilon$ with very small $\epsilon > 0$. In DA, TTC, and SD, the participant's ROL reverses the order of 1 and 2, if one of the following is true:*

- *the capacity of 2 is significantly larger compared to 1, but the options do not differ in terms of relative priority and popularity; or*
- *the participant's relative priority at 2 is significantly higher compared to 1, but the options do not differ in terms of capacity and popularity; or*
- *the perceived popularity of 2 is significantly lower compared to 1, but the options do not differ in terms of capacity and priority.*

The experiment on preference intensities by Klijn et al. (2013) provides evidence in line with this prediction. This observation is at odds with EBLA, since according to Meisner and von Wangenheim (2021) the profitability of this swap does not depend on $\mathbf{v}$.

Given a mechanism, an attainability distribution corresponds to beliefs about the reported preferences of other participants, but these beliefs do not have to be rational or even correct in any sense. In many settings, forming these beliefs correctly is complicated–even absent the usual biases in belief formation–because it is often unclear how the other side evaluates the proposers. Such aggregate uncertainty is persistent and does not vanish as markets grow large. This point is worth stressing as it raises the question of whether the classical 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. For example, a "hard-easy gap" (Dargnies et al., 2019) can be used to induce biased beliefs in experimental subjects if options evaluate participants according to a score in their own test. According to Prediction 3, such induced optimism or pessimism can trigger a swap deviation. Indeed, Rees-Jones and Skowronek (2018) find that overconfident participants tend play the truthful strategy more often, which is in line with our theory (but also with EBLA).

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. However, Proposition 2 not only holds in settings in which preferences are (believed to be) maximally negatively correlated such that each participant believes nobody else likes what she likes. The following prediction exploits the case in which (4) holds with equality as no participant knows enough to rank options according to attainability. The following prediction[9] is a corollary of Proposition 2.

Prediction 4. *Suppose all participants believe all ROLs and priority rankings of others are equally likely and that all options have the same capacity. Consider a participant who does not know her relative priority at any option. In DA or TTC, and for any* $\mathbf{v}$ *and any* ρ*, this participant ranks options according to* $\mathbf{v}$*, i.e., she submits the true ROL.*

Imagine all participants have preferences such that their v_s are individual iid draws and they all believe that options individually and privately draw priorities uniformly at random. Expecting that other participants are truthful implies that each ROL is submitted with the same probability, which together with the uniformly drawn priorities implies that all options are equally likely to be attainable so that (4) holds with equality for all participants. That is, in settings where preferen-

[9]The idea that a lack of information about others' preferences limits the benefits of strategic manipulations is not special to our model, see Roth and Rothblum (1999) or Coles and Shorrer (2014).

ces 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 Prediction 4 in spirit as they find that truthfulness rates in DA and TTC are highest when participants know nothing about the others' preferences (and priorities). 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. In the online appendix, we formulate a prediction tailored to their setting.

Prediction 4 can also serve to differentiate the effects of EBLA and report-dependent utility. Under EBLA, a participant never submits the true ROL if the attainability probability of her most-preferred option is sufficiently low, regardless of the attainability of other options, see Meisner and von Wangenheim (2021, Proposition 2). The true top choice is ranked down purely to avoid disappointment. In contrast, a participant with report-dependent utility needs a more attainable option to take its place in the ranking. In the setting of Prediction 4 such an option does not exist such that the true top choice is reported.

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. For any arbitrary ranking of acceptable options, we can construct beliefs such that this ROL is optimal even if report-dependent payoffs are arbitrarily small. 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. In the data and in line with our theory, truthfulness is indeed negatively associated with the perceived attainability of preferred options. More research is necessary to identify whether this trade-off between match quality and probability is preference-based or originates from misconceptions about the mechanism. This model leads to testable predictions to distinguish it from other approaches. Our insights also raise questions not answered in this paper, and we now briefly discuss some of these questions.

First, our decision-theoretic analysis does not consider strategic interaction and how equilibrium effects affect reporting behavior. However, our setting straightforwardly extends to multiple decision-makers, and it can be shown that preference misrepresentations persist in game-theoretic equilibrium. Suppose there are two participants whose private type consists of $\mathbf{v}$ and possibly a signal about relative priority such as test scores. Given a type distribution, participant 1 can compute an attainability distribution given her own type and a strategy, i.e., a mapping from types into ROLs, of participant 2. The optimal ROLs for all her types constitute her best response against the corresponding fixed strategy of player 2. If both players best-respond to each other, we have a Bayesian Nash equilibrium. In larger games, we can proceed in a similar fashion. In general, the existence of

a Bayesian Nash equilibrium is guaranteed by Milgrom and Weber (1985). It is easy to construct examples in which inefficiency (or justified envy) with respect to $\mathbf{v}$ persists in equilibrium allocations of TTC (or DA). That is, report dependence not only obstructs classical mechanisms from incentivizing truthful input, but also from implementing the allocations they are designed for.

There is a plethora of sources for a report-dependent payoff component, such as self-regarding concerns, aversion to rejections, or signaling motives in a larger game. The take-away message of this paper varies by context. First, market designers should be wary of factors that introduce report-dependent utility through the backdoor. For instance, changing the Turkish college admission mechanism to DA does not lead to a truthful dominant strategy if universities offer "preference scholarships," see Footnote 3. Second, report-dependent utility can also be generated by emotional factors, and under such an assumption growing evidence of non-truthful play in the field and in the lab can be explained. We thus caution against taking reported preferences at face value for policy decisions, and we emphasize the importance of participants' beliefs despite the strategy-proofness.

While we claim that our predictions in Section 3.1 are supported by experimental data, some readers might demand other evidence, and question whether the emotional motives play a relevant role in these experimental settings. In this context, our model may explain a preference for randomization observed in the field. Some participants deliberately introduce additional uncertainty which is inconsistent with standard preferences and, in particular, with EBLA which inherently entails an aversion to uncertainty. Each applicant in the German clearinghouse for university programs in medical fields has to submit three ROLs to three different procedures at the same time. Dwenger et al. (2018) document that applicants intentionally submit contradictory preferences and thereby essentially delegate their outcome to a suboptimal stochastic process. Through the lens of our model, applicants may prefer to delegate agency of their choice to mitigate emotional costs when the outcome differs from the reported preferences. Alternatively, contradicting ROLs can be used to justify ex-post that the outcome is in fact consistent with (one of the) reported preferences.

We did not investigate how to remedy the problems caused by non-truthful ROLs. This point immediately links to open empirical questions, aside from confirming our predictions. While we have argued for several plausible channels, our theory is silent on where the report dependence actually comes from. 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). In our model, 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. As an alternative, 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. To what extend avoiding negative

feedback about the own priority drives report-dependent utility is an empirical question. In the field, self-image protection does not seem to be the main driving factor. For instance, the self-selecting Mexican students studied by Chen and Pereyra (2019) know their own exam scores, and the cut-off scores for admission at each school are published after the match has finalized.

Recently, dynamic mechanisms have been suggested as a promising way to induce a truthful preference revelation. In settings with homogeneous priorities, 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). The uncertainty about attainability which can cause non-truthful reporting in static mechanisms can be reduced in dynamic mechanisms. For instance, iterative DA mechanisms (Bó and Hakimov, 2018, 2020a) or pick-an-object mechanisms (Bó and Hakimov, 2020b) do exactly that. First, they reveal if a favored option has already been selected by higher-priority agents, implying that unattainable options do not affect the choice. Second, they reveal if a favored option is still attainable, implying that ex-ante low attainability probabilities do not affect the choice. When participants only select from a pool of options left once it is their turn to choose, they can also credibly brag that they obtained their most-preferred option. The experimental evidence in favor of these mechanisms is in line with our theory. Under constraints on the possible preferences, stability or efficiency can be ensured despite report-dependent payoffs.

The attainability reduced form is not helpful when Lemma 2 does not apply. Immediate acceptance (IA, also known as the Boston mechanism) is a popular such (non-strategy-proof) mechanism. Here, the participant is not assigned to her highest-ranked attainable option. Therefore, match probabilities do not follow (2). For instance, consider student i_3 in the example in the appendix with IA and $R^{i_1} = (s_2, s_1, s_3)$. Submitting ROL (s_1, s_2, s_3) assigns i_3 to s_3. If she reported (s_2, s_1, s_3) instead, she would be accepted at s_2. So s_2 is available in the first step, but given i_1 and i_4 apply in the first step, s_2 has to reject i_3 in the second step. For this reason, utility changes due to a deviation to another ROL are more complicated to evaluate. As illustrated, the swap deviation from (s_1, s_2, s_3) to (s_2, s_1, s_3) can shift match probability from s_3 to s_2, which is impossible in a strategy-proof mechanism. This built-in feature of IA incentivizes ranking downwards competitive options and moving upwards safer options in the ranking. Report-dependent utility amplifies these incentives, but also gives rise to distinct incentives. Consequently, we can distinguish our theory from the idea that participants in DA submit manipulated ROLs because, for whatever reason, they think they play IA. For example, consider a participant in Li (2017) who has no report-dependent utility and who incorrectly assumes to play IA. For her, the commonly submitted ROL $(4, 3, 2, 1)$ is dominated by any other ROL not ranking 4 first.

However, Lemma 2 still applies when strategy-proof mechanisms are constrained by only allowing truncated ROLs. These mechanisms still assign our participant to her highest-ranked attainable option, but they impose that an outside option o (not getting matched) must have a rank $\widetilde{R}(o) \leq k$ for some $k < n$. Without report

dependence, this constraint provides incentives to rank better more attainable options to avoid falling back to option o. However, Lemma 2 ensures that all options in the optimal truncated ROL are ordered with respect to $\mathbf{v}$. Under report-dependent utility, the latter does not hold anymore for the same reasons as in the unconstrained mechanisms. Since the incentives to rank better more attainable options are already present in the unconstrained mechanism, truncated ROLs can be optimal under report-dependent utility. As the incentives go in the same direction, the welfare loss due to truncation constraints is expected to be lower under report-dependent utility. If each participant also has an endogenous outside option o' (such as a district school) and only prefers less than k options over it, the optimal ROL already satisfies the truncation constraint as all options ranked below o' are irrelevant.

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)[10] 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 report-independent 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. Similarly, advice suggesting truthful revelation may emphasize the report-independent dimension. Indeed, the "correct advice" in the same paper reads "The mechanism is designed so that truthful reporting maximizes your chances of getting favored schools. You should rank the schools in order of their true value to you." Here, the final sentence invokes a "true value," which attracts more attention to a payoff that is unrelated to a rank in the ROL. Thereby, the focus is shifted in a similar fashion.

Appendix

Attainability and examples

In any deterministic strategy-proof mechanism, we can employ the attainability distribution P as a reduced form summarizing beliefs about the other participants' ROLs, the options' priorities and their capacities, and we can use this distribution to calculate our participant's distribution over match outcomes for each ROL. For a given mechanism, the attainability state is fully determined by the other participants' ROLs, the options' priorities and capacities. Since the participant is always matched to the highest-ranked attainable option but cannot influence attainability herself, it is in her best interest to rank options according to $\mathbf{v}$ if report-independent utility is all she cares about.

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

Lemma 2. *A strategy-proof mechanism assigns a participant to her highest-ranked attainable option such that match probabilities are given by* (2).

Proof of Lemma 2. Consider any strategy-proof mechanism. Fix arbitrary ROLs of all other participants and let s be the highest-ranked attainable option in $\widetilde{R}$, our participant's ROL.

Suppose the participant is matched with s' ranked before s. But then, since s' is unattainable (i.e., she would not get in if ranked first), she would prefer $\widetilde{R}$ over her true ROL if s' was her most preferred option, a contradiction to strategy-proofness.

Suppose she is matched with s'' ranked behind s. But then, if $\widetilde{R}$ was the true ROL, she would prefer a match with s over s'', and ranking s first would achieve this match, again a contradiction to strategy-proofness. □

Let us consider Example 13.1 in Haeringer (2018), a school-choice setting with four students, $\{i_1, i_2, i_3, i_4\}$, and three schools, $\{s_1, s_2, s_3\}$. Schools s_1 and s_3 have unit capacity, while s_2 has two seats. We focus on the attainability state of student i_1 given the other participants submit the following ROLs:

$$R^{i_2} = (s_1, s_2, s_3), \quad R^{i_3} = (s_1, s_2, s_3), \quad R^{i_4} = (s_2, s_3, s_1),$$
$$R^{s_1} = (i_1, i_2, i_3, i_4), \quad R^{s_2} = (i_3, i_4, i_1, i_2), \quad R^{s_3} = (i_4, i_1, i_2, i_3),$$

where R^i is the ROL of agent i. We always mean the static implementation of the mechanisms, in which all participants simultaneously submit their ROLs in the beginning.

First, we consider student-proposing deferred acceptance (DA, with this abbreviation we always refer to the strategy-proof proposing side, not the possibly strategic receiving side). If i_1 submits $R^{i_1} = (s_2, s_1, s_3)$, she is assigned to s_1. Therefore, s_1 is attainable for her, but s_2 is not—she does not get in despite ranking it first. If, alternatively, i_1 ranked school s_3 first, she would get in: In step 1, only student i_3 gets rejected and moves on to school s_2 which accepts her, and the algorithm terminates. Consequently, s_3 is attainable as well. Second, we consider top-trading cycles (TTC). If i_1 ranks s_1 first, she forms a cycle with it in the first round. Similarly, she would be part of a (bigger) cycle in the first round if she ranked s_2 or s_3 first. Therefore, each school is attainable. Third, we consider serial dictatorship (SD) in the order (i_4, i_3, i_2, i_1). In this setting, i_1 gets to choose last and is matched to the remaining option, s_3, for every possible ROL. Table 2 summarizes this analysis by stating the attainability state for each mechanism for the given ROLs.

The attainability states in Table 2 correspond to fixed ROLs of the other participants and capacities. An attainability distribution P is a probability distribution over such states, and it corresponds to a probability distribution over the other ROLs (and possibly capacities) in the same fashion as above. Attainability states are usually not independent even when all participants' preferences are independently distributed. For instance, if each of our exemplary participants

Mechanism	Attainability state		
	A_{s_1}	A_{s_2}	A_{s_3}
DA	1	0	1
TTC	1	1	1
SD (last chooser)	0	0	1

Table 2: The attainability states for each of the three discussed mechanisms and the given ROLs and priorities.

independently draws v_{s_1} from a uniform distribution on $[0.5, 1.5]$, but v_{s_2} and v_{s_3} independently from a uniform distribution on $[0, 1]$, school s_1 is more in demand. If then additionally all three participants independently draw an individual priority score which determines the priority order at all options, attainability of s_1 makes attainability of s_2 or s_3 more likely.

The three mechanisms above are deterministic: given all ROLs the allocation is certain. The attainability reduced form can also incorporate stochastic mechanisms. For instance, consider a lottery that independently of the ROL randomly assigns student i_1 to s_1 and s_3 with probability $1/4$, each, and to s_2 with probability $1/2$. Trivially, the true ROL is always optimal here because the ROL does not affect the allocation. We can extend our setting such that in addition to the ROLs of all other players, a chance player ("Nature") influences the attainability state. In this example, the attainability distribution simply assigns probabilities $1/4, 1/4$, and $1/2$ to attainability states $(1, 0, 0)$, $(0, 0, 1)$, and $(0, 1, 0)$, respectively. Such a chance player can also represent tie-breaking when priorities are weak, or randomize over the order of choice in SD. In contrast, attainability is not a helpful concept when Lemma 2 does not apply. With Immediate acceptance, we discuss such a non-strategy-proof example in Section 4.

Proofs

Proof of Proposition 1. Fix any arbitrary ROL $\widetilde{R} = (\widetilde{1}, \widetilde{2}, \ldots, \widetilde{n})$, any function ρ, and any report-independent utility vector $\mathbf{v}$. We construct an attainability distribution $\widetilde{P}$ 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. This is just a normalization of the payoffs given that the participant can improve upon any ROL listing an unacceptable option with $v_{n+1} < v_n$ by simply dropping it behind the outside option.

The constructed $\widetilde{P}$ 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

$$q_{\widetilde{r}} > q_{\widetilde{r+1}} \quad \forall r \leq \widetilde{R}(n) \tag{8}$$

with $\sum_{r=1}^{\widetilde{R}(n)} q_{\widetilde{r}} = 1$. 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

$$U_\rho(\mathbf{v}|\widetilde{R}) - U_\rho(\mathbf{v}|\widehat{R}) \geq \sum_{r=1}^{\widetilde{R}(n)} q_{\widetilde{r}}\big(\rho(r) - \rho(\widehat{R}(\widetilde{r}))\big), \tag{9}$$

because in each state both ROLs either yield the same report-independent utility $v_{\widetilde{r}}$ or $\widehat{R}$ yields $v_n = 0 < v_{\widetilde{r}}$ such that we can restrict attention to comparing report-dependent utility. Since ρ is decreasing and (8) holds, $\widetilde{R}$ puts the largest $\rho(r)$ on the most likely states. Hence, (9) 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 $\widetilde{P}$, which can only decrease report-dependent utility.

Next, we compare $\widetilde{R}$ to truncations of itself. Suppose $\widehat{R}(\widetilde{r}) = \widetilde{R}(\widetilde{r})$ 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

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

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

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

which is positive for all t if

$$q_{\widetilde{t}} \geq \left(1 - \sum_{r=1}^{t-1} q_{\widetilde{r}}\right)\frac{\rho(t) - \rho(\widetilde{R}(n))}{v_{\widetilde{t}} + \rho(t) - \rho(\widetilde{R}(n))} = \left(1 - \sum_{r=1}^{t-1} q_{\widetilde{r}}\right)\alpha \quad \forall t < \widetilde{R}(n), \tag{10}$$

where $\alpha \in (0, 1)$ because $\rho(t) > \rho(\widetilde{R}(n))$ for all $t < \widetilde{R}(n)$. If additionally (8) holds, also all other truncated ROLs of length t yield a lower expected payoff than $\widetilde{R}$.

Because $\widetilde{R}$ is a strict utility maximizer given the distribution $\widetilde{P}$ constructed above and expected utility is continuous in P, we can construct an open ball around $\widetilde{P}$ such that both (8) and (10) hold for all P in this open ball. □

Proof of Proposition 2. First, note that

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

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 (4) is violated for some $\overline{r}$ of ROL $\widetilde{R}$, and let the difference in (4) 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

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

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

Suppose (4) holds for all $\overline{r}$, and fix any arbitrary $\mathbf{v}$ and ρ. Under strategy-proofness, $U_\rho(\mathbf{v}|R) - U_\rho(\mathbf{v}|\widetilde{R}) \geq \Delta_\rho$, and we see that (4) 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 (4) holds. □

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

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

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

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 (5),

$$\sum_{r=k}^{\ell-1} \Big((f_r - f_\ell)\Delta_r + \delta_r^{R,\widetilde{R}} (\rho(k) - \rho(r+1)) \Big) < \sum_{r=k}^{\ell-1} \delta_r^{R,\widetilde{R}} (v_r - v_\ell).$$

□

References

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

Bó, I., Hakimov, R., 2018. The iterative deferred acceptance mechanism. Tech. rep.

Bó, I., Hakimov, R., 2020a. Iterative versus standard deferred acceptance: Experimental evidence. The Economic Journal 130 (626), 356–392.

Bó, I., Hakimov, R., 2020b. Pick-an-object mechanisms. Available at SSRN.

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.

Coles, P., Shorrer, R., 2014. Optimal truncation in matching markets. Games and Economic Behavior 87, 591–615.

Dargnies, M.-P., Hakimov, R., Kübler, D., 2019. Self-confidence and unraveling in matching markets. Management Science 65 (12), 5603–5618.

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.

Dwenger, N., Kübler, D., Weizsäcker, G., 2018. Flipping a coin: Evidence from university applications. Journal of Public Economics 167, 240–250.

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

Fernandez, M. A., Rudov, K., Yariv, L., 2022. Centralized matching with incomplete information. American Economic Review: Insights 4 (1), 18–33.

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.

Haeringer, G., 2018. Market design: auctions and matching. MIT Press.

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.

Klijn, F., Pais, J., Vorsatz, M., 2013. Preference intensities and risk aversion in school choice: A laboratory experiment. Experimental Economics 16 (1), 1–22.

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.

Milgrom, P. R., Weber, R. J., 1985. Distributional strategies for games with incomplete information. Mathematics of Operations Research 10 (4), 619–632.

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.

Roth, A. E., Rothblum, U. G., 1999. Truncation strategies in matching markets—in search of advice for participants. Econometrica 67 (1), 21–43.

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