Strategy-proof Pareto-improvement*

Samson Alva
University of Texas at San Antonio
samson.alva@gmail.com

Vikram Manjunath
University of Ottawa
vikramma@gmail.com

September 17, 2018

Abstract

We consider a model where each agent has an outside option of privately known value. We show that one strategy-proof and individually rational mechanism Pareto-improves another if and only if it expands the set of participants. Corollaries include: a sufficient condition for a mechanism to be on the Pareto-efficient frontier of strategy-proof mechanisms; uniqueness of strategy-proof Pareto-improvements under true preferences over certain normatively meaningful benchmark allocation rules; and a characterization of the pivotal mechanism.

Keywords: strategy-proofness, Pareto-improvement, Pareto-constrained participation-maximality, school choice, pivotal mechanism
JEL Codes: C78; D47; D71; D82

1 Introduction

We consider mechanisms that choose an allocation as a function of agents' preferences. In our model, each allocation is associated with a fixed set of participants.[1] A

*We thank Inácio Bó, Alexander Brown, Rossella Calvi, Lars Ehlers, Ben Golub, Guillaume Haeringer, Patrick Harless, Eun Jeong Heo, Sean Horan, Fuhito Kojima, Scott Kominers, Silvana Krasteva, Timo Mennle, Debasis Mishra, Thayer Morrill, Mallesh Pai, William Phan, Santanu Roy, Erling Skancke, Tayfun Sönmez, Alexander Teytelboym, William Thomson, Guoqiang Tian, Ryan Tierney, Juuso Toikka, Utku Ünver, Rodrigo Velez, Dan Walton, Alexander Westkamp, Alexander Wolitzky, and Bumin Yenmez for helpful comments and discussions. We also thank seminar audiences at the University of Rochester, the University of Cologne, the University of Bonn, Southern Methodist University, the University of Windsor, Harvard/MIT, the University of Ottawa, North Carolina State University, Carleton University, Université de Montréal, and participants at numerous conferences for their feedback. We are also grateful for the helpful suggestions of the editor and two anonymous referees.

[1] Instances of the model include object allocation (with priorities) [Hylland and Zeckhauser, 1979, Abdulkadiroğlu and Sönmez, 2003], matching with contracts [Hatfield and Milgrom, 2005], excludable public goods [Jackson and Nicolò, 2004], and more.

non-participant at an allocation consumes his outside option. A mechanism is strategy-proof if, at each profile of preferences, no agent is able to beneficially misreport his preferences. It is individually rational if, at each profile of preferences, no participant prefers his outside option to the chosen allocation.

We show the equivalence of two binary relations over strategy-proof and individually rational mechanisms. The first is the Pareto-improvement relation: mechanism φ (weakly) Pareto-improves mechanism φ' if, at each profile of preferences, each agent finds the choice of φ at least as desirable as the choice of φ'. The other is the participation-expansion relation: φ (weakly) participation-expands φ' if, at each profile of preferences, each participant at the choice of φ' is a participant at the choice of φ. Therefore, the requirements of strategy-proofness and individual rationality contain enough information about preferences to ensure that a comparison that makes no reference to preferences (participation-expansion) is equivalent to one that does (Pareto-improvement).[2]

We make three assumptions on preferences. Firstly, we rule out externalities on non-participants. So, if an agent participates at neither allocation α nor at allocation β, then he is indifferent between α and β. Secondly, only to show that participation-expansion implies Pareto-improvement, we assume the range of values that an agent's outside option may take are on the order of what the mechanism may offer him as a participant. We call this second assumption "richness of the outside option." It ensures that the (upward) movement of an agent's outside option in his preference is essentially unrestricted. Since how an agent compares his outside option to the various alternatives as a participant is often private information, this is natural in many applications. Finally, only to show that Pareto-improvement implies participation-expansion, we assume that no agent is indifferent between his outside option and any allocation at which he participates. We call this final assumption "no indifference with the outside option."

We point out three useful corollaries of our main result:

1. Say that a pair of mechanisms, φ and φ', are participation-equivalent if, for each profile of preferences, the allocation chosen by φ has the same participants as the allocation chosen by φ'. *Given no externalities on non-participants and richness of the outside option, if a pair of individually rational and strategy-proof mechanisms are participation-equivalent, then they are also welfare-equivalent.* A further corollary is a simple, yet novel, characterization of the pivotal mechanism for the problem of strategy-proof social choice with continuous transfers.

[2] To our knowledge, the only other result relating the Pareto-improvement relation to participation is for the probabilistic allocation of objects: strict Pareto-improvement implies strict participation-expansion [Erdil, 2014].

2. Say that an allocation is Pareto-constrained participation-maximal if every allocation that strictly expands the set of participants makes some agent worse off. Say that a mechanism is strategy-proofness-constrained Pareto-efficient if no strategy-proof mechanism strictly Pareto-improves it. *Given our three assumptions, a sufficient condition for a strategy-proof and individually rational mechanism to be strategy-proofness-constrained Pareto-efficient is that it always selects a Pareto-constrained participation-maximal allocation.* This implies the strategy-proofness-constrained Pareto-efficiency of the student-optimal stable mechanism for school choice problems and of the cumulative offer mechanism for many matching problems with complex constraints.[3]

3. Suppose a normative rule associates each profile of preferences with some Pareto-constrained participation-maximal and individually rational allocation. *Given our three assumptions, there is at most one (in welfare terms) strategy-proof mechanism that selects, for each profile of preferences, an allocation that each agent finds at least as desirable as that prescribed by the benchmark under true preferences.*[4] This result tells us that the student-optimal stable mechanism is the only strategy-proof mechanism satisfying various normative requirements that are weaker than stability.[5]

The rest of the paper is organized as follows. We introduce our model in Section 2. We define properties of allocations and mechanisms in Section 3. We state and prove our results in Section 4. We discuss applications in Section 5. We defer detailed discussion of related literature to Sections 4 and 5.

2 The Model

Let N be a finite set of **agents**. Let $\mathcal{F}$ be the nonempty set of **allocations**. Given $\alpha \in \mathcal{F}$, let $\boldsymbol{N(\alpha)} \subseteq N$ be the **participants** at α.[6] For each $i \in N$, let $\boldsymbol{\mathcal{F}_i} \subseteq \mathcal{F}$ be the set of allocations that i participates at.[7] If $\alpha \in \mathcal{F}$ is chosen and i is not a participant at α, then i consumes his **outside option**. We denote consumption of the outside option by $\boldsymbol{\varnothing}$. For each $\alpha \in \mathcal{F}$, we denote by $\boldsymbol{\alpha(i)}$ either α if $\alpha \in \mathcal{F}_i$ or $\varnothing$ otherwise. Note that the set of participants,

[3] See Hirata and Kasuya [2017] for another sufficient condition.

[4] However, this does not hold if the benchmark rule identifies allocations that are not individually rational or Pareto-constrained participation-maximal, as would be the case when the benchmark rule is a constant function, identifying a single benchmark *allocation* independent of preferences.

[5] These include "legality" [Morrill, 2016] and partial fairness [Dur et al., 2015].

[6] For instance, in the assignment of indivisible objects to agents, participants are those agents who receive an object.

[7] This set is define by $\mathcal{F}_i = \{\alpha \in \mathcal{F} : i \in N(\alpha)\}$.

$N(\alpha)$, at each α is a part of the specification of the model. That is, the definition of α *presumes* the participation of $N(\alpha)$. In the next section, we define *individual rationality* to account for agents' preferences in regards to participation.

For each $i \in N$, his preference is a complete, reflexive, and transitive binary relation on $\mathcal{F}_i \cup \{\emptyset\}$. We denote it by $\boldsymbol{R_i}$. Since i's preference is over $\mathcal{F}_i \cup \{\emptyset\}$ rather than $\mathcal{F}$, we have assumed that i is indifferent between any pair of allocations that he does not participate at. Consequently, his welfare from such allocations is fully determined by his outside option. In effect, we assume **no externalities on non-participants**.

For each pair $\alpha, \beta \in \mathcal{F}$, we write $\boldsymbol{\alpha(i)\ R_i\ \beta(i)}$ to mean that i finds $\alpha(i)$ to be at least as good as $\beta(i)$. We use $\boldsymbol{P_i}$ to denote strict preference and $\boldsymbol{I_i}$ to denote indifference, the asymmetric and symmetric components of R_i respectively. Let $\boldsymbol{\mathcal{R}_i}$ be a set of preference relations for i. A *preference domain* is $\boldsymbol{\mathcal{R}} \equiv \times_{i \in N} \mathcal{R}_i$.

Our analysis is for fixed N, $\mathcal{F}$, and $\mathcal{R}$. Thus, an economy is entirely described by $R \in \mathcal{R}$. A (direct) **mechanism**, $\varphi : \mathcal{R} \to \mathcal{F}$, associates each economy with an allocation. For each $R \in \mathcal{R}$ and each $i \in N$, instead of $\varphi(R)(i)$, we write $\boldsymbol{\varphi_i(R)}$.

3 Properties of Allocations and Mechanisms

Individual rationality An allocation is individually rational if each agent finds it at least as desirable as not participating. That is, for each $R \in \mathcal{R}$ and each $\alpha \in \mathcal{F}$, α is individually rational at R if, for each $i \in N$, $\alpha(i)\ R_i\ \emptyset$.[8] A mechanism, φ, is *individually rational* if, for each $R \in \mathcal{R}$, $\varphi(R)$ is individually rational at R. Individual rationality ensures that no agent has an incentive to exercise his outside option when the allocation chosen by the mechanism relies on his presence.

Participation-expansion An allocation participation-expands another if participation at the latter entails participation at the former. That is, for each pair $\alpha, \beta \in \mathcal{F}$, α participation-expands β if $N(\alpha) \supseteq N(\beta)$, and *strictly* so if $N(\alpha) \supsetneq N(\beta)$. If they have the same participants, then they are **participation-equivalent**. That is, α is participation-equivalent to β if $N(\alpha) = N(\beta)$. Given a pair of mechanisms φ and φ', φ *participation-expands* φ' if, for each $R \in \mathcal{R}$, $\varphi(R)$ participation-expands $\varphi'(R)$. They are *participation-equivalent* if, for each $R \in \mathcal{R}$, $\varphi(R)$ and $\varphi'(R)$ are participation-equivalent.

Pareto-improvement One allocation Pareto-improves another if each agent finds the first at least as desirable as the second. That is, for each $R \in \mathcal{R}$ and each pair $\alpha, \beta \in \mathcal{F}$,

[8] Since each $i \notin N(\alpha)$ consumes $\emptyset$, it suffices to verify that for each $i \in N(\alpha), \alpha(i)\ R_i\ \emptyset$.

α Pareto-improves β at R if, for each $i \in N$, $\alpha(i)\ R_i\ \beta(i)$.[9] If α Pareto-improves β at R and there is $i \in N$ such that $\alpha(i)\ P_i\ \beta(i)$, then α **strictly Pareto-improves** β at R. If $\alpha \in \mathcal{F}$ is such that no allocation strictly Pareto-improves it at R, then α is **Pareto-efficient** at R. For each pair of mechanisms, φ and φ', φ *Pareto-improves* φ' if, for each $R \in \mathcal{R}$, $\varphi(R)$ Pareto-improves $\varphi'(R)$ at R. If φ Pareto-improves φ' and for some $R \in \mathcal{R}$, $\varphi(R)$ strictly Pareto-improves $\varphi'(R)$ at R, then φ *strictly Pareto-improves* φ'. If, for each $R \in \mathcal{R}$ and each $i \in N$, $\varphi_i(R)\ I_i\ \varphi'_i(R)$, then φ and φ' are **welfare-equivalent**. If, for each $R \in \mathcal{R}$, $\varphi(R)$ is Pareto-efficient at R, then φ is *Pareto-efficient*.

Pareto-constrained participation-maximality An allocation is Pareto-constrained participation-maximal at a given $R \in \mathcal{R}$ if there is no other allocation that strictly expands the set of participants without harming anyone. That is, for each $R \in \mathcal{R}$ and each $\alpha \in \mathcal{F}$, α is Pareto-constrained participation-maximal at R if there is no $\beta \in \mathcal{F}$ such that (*1*) $N(\alpha) \subsetneq N(\beta)$, so that at least one additional agent participates at β compared to α and (*2*) there is no $i \in N$, such that $\alpha(i)\ P_i\ \beta(i)$, so that nobody is worse off at β than at α. Equivalently, an allocation is Pareto-constrained participation-maximal if there is no other allocation that simultaneously Pareto-improves it and strictly participation-expands it.[10] By definition, each allocation that is individually rational and Pareto-efficient is Pareto-constrained participation-maximal, but the converse is not true.[11] A mechanism, φ, is *Pareto-constrained participation-maximal* if, for each $R \in \mathcal{R}$, $\varphi(R)$ is Pareto-constrained participation-maximal at R.

IR-PCPM-Pareto-connectedness The Pareto-improvement relation is reflexive and transitive but not complete. Two allocations are **Pareto-comparable** if one Pareto-improves the other. Two individually rational and Pareto-constrained participation-maximal allocations are IR-PCPM-Pareto-connected if there is a sequence of individually rational and Pareto-constrained participation-maximal allocations starting at one and ending at the other such that successive allocations are Pareto-comparable. That is, two individually rational and Pareto-constrained participation-maximal allocations, $\alpha, \beta \in \mathcal{F}$, are IR-PCPM-Pareto-connected if there is a sequence $(\alpha_k)_{k=0}^{K}$, with $\alpha_0 \equiv \alpha$ and $\alpha_K \equiv \beta$, such that for each $k \in \{1, \ldots, K\}$, α_k is individually rational, Pareto-constrained

[9] In this case, some authors say that α *weakly* Pareto-improves β. However, since this is the main form of Pareto-improvement that we consider, we drop the qualifier.

[10] We have chosen this terminology since an allocation satisfies this property when its set of participants is maximal subject to the constraint of Pareto-improvement.

[11] In the special case of the object allocation model with strict preferences (see Section 5.1), where non-wastefulness is defined, non-wastefulness is stronger than Pareto-constrained participation-maximality.

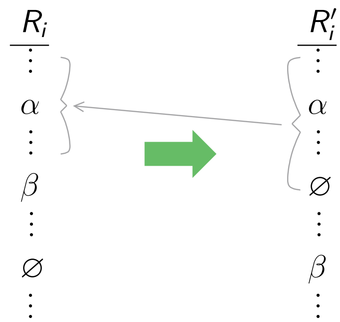

Figure 1: Suppose there is a preference relation, R_i, where i prefers α to β and finds β to be at least desirable as his outside option. Richness of the outside option says that there is another preference relation for i, R_i', where α is better than his outside option, which is in turn better than β. Furthermore, any allocation that i finds at least as desirable as his outside option under R_i', he prefers over β under R_i.

participation-maximal, and Pareto-comparable to α_{k-1}. Each pair of individually rational and Pareto-constrained participation-maximal mechanisms, φ and φ', are *IR-PCPM-Pareto-connected* if, for each $R \in \mathcal{R}$, $\varphi(R)$ and $\varphi'(R)$ are IR-PCPM-Pareto-connected at R.

Strategy-proofness A mechanism, φ, is strategy-proof if no agent can benefit by misreporting his preferences, no matter what other agents do. That is, for each $R \in \mathcal{R}$, each $i \in N$, and each $R_i' \in \mathcal{R}_i$, $\varphi_i(R)\ R_i\ \varphi_i(R_i', R_{-i})$.

Strategy-proofness-constrained Pareto-efficiency A strategy-proof mechanism is strategy-proofness-constrained Pareto-efficient if there is no strategy-proof mechanism that *strictly* Pareto-improves it.

4 Strategy-proof Pareto-improvement

We present here our main result and several corollaries, which we show under varying combinations of two further assumptions on preferences.

The first is **richness of the outside option** (Figure 1): for each $i \in N$, each $R_i \in \mathcal{R}_i$, and each pair $\alpha, \beta \in \mathcal{F}_i$ such that $\alpha(i)\ P_i\ \beta(i)\ R_i\ \varnothing$, there is $R_i' \in \mathcal{R}_i$ such that (1) $\alpha(i)\ P_i'\ \varnothing\ P_i'\ \beta(i)$, and (2) for each $\gamma \in \mathcal{F}_i$, if $\gamma(i)\ R_i'\ \varnothing$ then $\gamma(i)\ P_i\ \beta(i)$. The second assumption is **no indifference with the outside option**: for each $i \in N$ and each $R_i \in \mathcal{R}_i$, there is no $\alpha \in \mathcal{F}_i$ such that $\alpha(i)\ I_i\ \varnothing$.[12]

[12] Sönmez [1999] makes two similar assumptions, where an agent's endowment plays the role that $\varnothing$ plays here. Also see Erdil and Ergin [2017] for an instance of no indifference with the outside option in a matching setting with weak preferences.

Remark 1. In many discrete private goods applications, it is typical to assume an agent's preference domain consists of all strict preference relations over own outcomes. This, however, is *far* more than what is implied by our assumptions. No indifference with the outside option only states that ties with $\varnothing$ are broken, but other indifferences may exist. More importantly, richness of the outside option is a great deal weaker than the assumption that all orderings are available. To see this, consider the fact that R_i and R_i' in the definition of richness need not preserve, entirely, the relative orderings of alternatives other than α, β, and $\varnothing$. This makes it even weaker than the analogous assumption in Sönmez [1999]. We discuss the necessity of these assumptions in the Online Appendix.

Our main result relates the Pareto-improvement and participation-expansion relations over strategy-proof and individually rational mechanisms. Under each of the above assumptions, we show that one of these relations is a refinement of the other. Thus, under both assumptions together, they are equivalent.

Theorem 1. *Consider the Pareto-improvement and the participation-expansion relations restricted to the set of strategy-proof and individually rational mechanisms.*

(A) If the preference domain satisfies no indifference with the outside option, then Pareto-improvement implies participation-expansion.

(B) If the preference domain satisfies richness of the outside option, then participation-expansion implies Pareto-improvement.

Proof.

(A) Let φ and φ' be a pair of individually rational mechanisms such that φ' Pareto-improves φ. Then, for each $R \in \mathcal{R}$, and each $i \in N$, $\varphi_i'(R)\ R_i\ \varphi_i(R)$. By individual rationality of φ and no indifference with the outside option, if $i \in N(\varphi(R))$, then $\varphi_i(R)\ P_i\ \varnothing$, so $\varphi_i'(R)\ P_i\ \varnothing$. Thus, $i \in N(\varphi'(R))$. Since this holds at each R for each $i \in N(\varphi(R))$, φ' participation-expands φ.

(B) Let φ and φ' be a pair of strategy-proof and individually rational mechanisms such that φ' participation-expands φ. If φ' does not Pareto-improve φ, then there are $i \in N$ and $R \in \mathcal{R}$ such that $\varphi_i(R)\ P_i\ \varphi_i'(R)$. Let $\alpha \equiv \varphi(R)$ and $\beta \equiv \varphi'(R)$. Since φ' participation-expands φ, $N(\beta) \supseteq N(\alpha)$. Since both φ and φ' are individually rational, $\alpha(i)\ R_i\ \varnothing$ and $\beta(i)\ R_i\ \varnothing$. Since $\alpha(i)\ P_i\ \beta(i)\ R_i\ \varnothing$, we deduce that $i \in N(\alpha)$. Thus, since $N(\beta) \supseteq N(\alpha)$, $i \in N(\beta)$.

Since $\alpha(i)\ P_i\ \beta(i)\ R_i\ \varnothing$, by richness of the outside option, there is $R_i' \in \mathcal{R}_i$ such that (1) $\alpha(i)\ P_i'\ \varnothing\ P_i'\ \beta(i)$, and (2) for each $\gamma \in \mathcal{F}_i$, if $\gamma(i)\ R_i'\ \varnothing$ then $\gamma(i)\ P_i\ \beta(i)$. Let $\gamma \equiv \varphi(R_i', R_{-i})$ and $\gamma' \equiv \varphi(R_i', R_{-i})$.

Since φ is strategy-proof, $\gamma(i)\ R_i'\ \alpha(i)$. Otherwise, i would have an incentive to misreport R_i if his true preferences are R_i'. Thus, by definition of R_i', $\gamma(i)\ P_i'\ \varnothing$. So $i \in N(\gamma)$. Again, since φ' participation-expands φ, $i \in N(\gamma')$. Since φ' is individually rational, $\gamma'(i)\ R_i'\ \varnothing$. By definition of R_i', $\gamma'(i)\ P_i\ \beta(i)$. This contradicts the strategy-proofness of φ' since i has an incentive to misreport his preference as R_i' if his true preferences are R_i. From this contradiction, we conclude that φ' does Pareto-improve φ.

□

For individually rational mechanisms, Theorem 1(A) follows from the assumption of no indifference with the outside option and the definition of individual rationality. In fact, it does not even require the mechanisms being compared to be strategy-proof. The more novel and compelling aspect of Theorem 1 is Part (B). It says that for a pair of mechanisms that are both strategy-proof and individually rational, participation-expansion implies Pareto-improvement even if agents may be indifferent between participating and not participating, given richness of the outside option.

Remark 2. The proof of Theorem 1 is at the level of a single agent. Consequently, if strategic preference reporting is restricted to only a subset S of agents, a version of the theorem for S holds: if the preferences of agents in S satisfy richness of the outside option and no indifference with the outside option, then among mechanisms that are strategy-proof and individually rational for members of S, the versions of the Pareto-improvement and participation-inclusion relations restricted to S coincide.

For the probabilistic allocation of indivisible goods when each agent has unit demand and strict preferences, Erdil [2014] has shown that strict Pareto-improvement in the stochastic dominance sense implies a strict participation-expansion. However, we are heretofore unaware of other work showing an *equivalence* between the Pareto-improvement and participation-expansion relations.

Theorem 1 provides an understanding of the Pareto-improvement relation over the set of strategy-proof and individually rational mechanisms. The implications of this result, some of which we consider in the remainder of this section, have many applications that we describe in Section 5.

By selecting an allocation for each profile of preferences, a mechanism selects a list of participants for each profile of preferences as well—those agents who participate at the allocation that it selects. The following corollary of Theorem 1(B) says that, under richness of the outside option, pinning down who participates at each profile pins down the entire mechanism in welfare terms. It does not require $\mathcal{R}$ to satisfy no indifference with the outside option, and so is relevant for economies with continuous preferences, as in the case where continuous transfers are possible.

Corollary 1. (Participation-equivalence) *Let the preference domain satisfy richness of the outside option. If a pair of strategy-proof and individually rational mechanisms are participation-equivalent, then they are welfare-equivalent.*

Under the assumption of no indifference with the outside option, the set of individually rational and Pareto-constrained participation-maximal allocations has a particularly nice structure. The IR-PCPM-Pareto-connectedness relation over this set is reflexive, symmetric, and transitive, so it is an equivalence relation. Therefore, it partitions this set into components that are IR-PCPM-Pareto-connected. The following lemma says that every allocation in the same component involves the same participants. Since the statement of the lemma is specific to a fixed profile of preferences, it does not rely on richness of the outside option.[13]

Lemma 1 (Structure Lemma)**.** *Let the preference domain satisfy no indifference with the outside option. For each profile of preferences, if a pair of individually rational and Pareto-constrained participation-maximal allocations are IR-PCPM-Pareto-connected, then they are also participation-equivalent.*

Proof. Let $R \in \mathcal{R}$ and $\alpha \in \mathcal{F}$ such that α is individually rational and Pareto-constrained participation-maximal at R. We first show for each $\beta \in \mathcal{F}$, if β Pareto-improves α at R, then $N(\alpha) = N(\beta)$ and β is individually rational and Pareto-constrained participation-maximal at R as well. Then we show for each $\beta \in \mathcal{F}$, if β is individually rational, Pareto-constrained participation-maximal, and IR-PCPM-Pareto-connected to α, then α and β are participation-equivalent.

For each $i \in N(\alpha)$, by no indifference with the outside option, $\alpha(i)\ P_i\ \varnothing$. Since $\beta(i)\ R_i\ \alpha(i)$, $\beta(i)\ P_i\ \varnothing$, and so $i \in N(\beta)$, and β is individually rational at R. Thus, $N(\alpha) \subseteq N(\beta)$. Since β Pareto-improves α, Pareto-constrained participation-maximality of α implies it is not the case that $N(\beta) \supsetneq N(\alpha)$. Then together with $N(\alpha) \subseteq N(\beta)$, we have $N(\beta) = N(\alpha)$.

Let $\gamma \in \mathcal{F}$. If γ Pareto-improves β, then it Pareto-improves α. Since α is Pareto-constrained participation-maximal, it is not the case that $N(\gamma) \supsetneq N(\alpha)$. However,

[13] In Alva and Manjunath [2017], we draw stronger conclusions by adding structure to the model. Indeed, the Structure Lemma is closely related to the Rural Hospitals Theorem in two-sided matching [Roth, 1986].

$N(\alpha) = N(\beta)$, so it is not the case that $N(\gamma) \supsetneq N(\beta)$. Since this holds for any γ that Pareto-improves β, β is Pareto-constrained participation-maximal at R.

Now, suppose β is individually rational, Pareto-constrained participation-maximal, and α and β are IR-PCPM-Pareto-connected. Then, there is a sequence of individually rational and Pareto-constrained participation-maximal allocations, $\left(\alpha^k\right)_{k=0}^K$, with $\alpha^0 = \alpha$ and $\alpha^K = \beta$, such that for each $k \in \{1, \dots, K\}$, either α^k Pareto-improves α^{k-1} or vice versa. In either case, by the argument above, $N(\alpha^k) = N(\alpha^{k-1})$, and so $N(\alpha) = N(\beta)$. □

A consequence of Corollary 1 and the Structure Lemma is that no two welfare-distinct strategy-proof mechanisms select from the same component of the partition of the individually rational and Pareto-constrained participation-maximal allocations that we have described above at every preference profile.

Proposition 1. *Let the preference domain satisfy richness of the outside option and no indifference with the outside option. If a pair of strategy-proof, individually rational, and Pareto-constrained participation-maximal mechanisms are IR-PCPM-Pareto-connected, then they are welfare-equivalent.*

Proof. Let φ and φ' be a pair of strategy-proof, individually rational, and Pareto-constrained participation-maximal mechanisms. If they are IR-PCPM-Pareto-connected, then by the Structure Lemma, they are participation-equivalent. Thus, by the Corollary 1, they are welfare-equivalent. □

Previous work on describing the Pareto frontier of strategy-proof mechanisms includes Anno and Kurino [2016] on object allocation with multi-unit demand and no transfers, Ohseto [2006] and Sprumont [2013] on object allocation with unit demand and transfers, and Anno and Sasaki [2013] on divisible goods allocation. Proposition 1 contributes by providing a useful sufficient condition for a mechanism to be on this frontier.

Corollary 2. *Let the preference domain satisfy richness of the outside option and no indifference with the outside option. A strategy-proof and individually rational mechanism is strategy-proofness-constrained Pareto-efficient if it is Pareto-constrained participation-maximal.*[14]

A natural context where one might consider IR-PCPM-Pareto-connected mechanisms is when there is a normative benchmark rule that identifies a particular individually rational and Pareto-constrained participation-maximal allocation for every profile of preferences. If a designer is tasked with choosing a strategy-proof mechanism that Pareto-

[14] However, Pareto-constrained participation-maximality is not a necessary condition. See the Online Appendix for a counterexample.

improves on such a benchmark rule under agents' true preferences, Proposition 1 says that his problem has at most one solution.

Corollary 3. *Let the preference domain satisfy richness of the outside option and no indifference with the outside option. Consider an individually rational and Pareto-constrained participation-maximal benchmark rule. In welfare terms, there is at most one strategy-proof mechanism that Pareto-improves it under true preferences.*

The limitation of the above corollary is that it requires the benchmark to be individually rational and Pareto-constrained participation-maximal. Without these qualifications, the conclusion does not hold. Suppose, as is the case in many applications, it is possible to assign every agent his outside option. Then every individually rational mechanism Pareto-improves the benchmark that always assigns the outside option to every agent. Of course, there may be many strategy-proof and individually rational mechanisms. Yet, this does not contradict Corollary 3 since this benchmark is not Pareto-constrained participation-maximal.

5 Applications

We consider several applications of our results: the object allocation model, augmented with choice correspondences (Section 5.1); school choice (Section 5.2); excludable public goods (Section 5.3); and transferable utility (Section 5.4).

5.1 Object Allocation and Matching

We start by explaining how the object allocation model fits into the general model that we defined in Section 2.

The object allocation model Let $\boldsymbol{O}$ be a finite and nonempty set of objects, $\boldsymbol{T}$ be a nonempty set of terms under which an agent may be assigned an object, and $\boldsymbol{X} \subseteq N \times O \times T$ be a nonempty set of possible triples. The triple $(i, o, t) \in N \times O \times T$ represents "i consumes o under the terms t." These are *contracts* in Hatfield and Milgrom [2005]. For each $x \in X$, let $\boldsymbol{N(x)}$ be the agent associated with x. For each $Y \subseteq X$, let $\boldsymbol{N(Y)}$ be the set of agents associated with triples in Y. For each $i \in N$, let $\boldsymbol{Y(i)}$ be the triples in Y associated with i. For each $o \in O$, let $\boldsymbol{Y(o)}$ be the triples in Y associated with o. Each object may only be allocated in certain ways. These constraints define, for each $o \in O$, the **feasible sets** for o, which is a collection of subsets of $X(o)$. We denote it by $\boldsymbol{\mathcal{F}_o}$. In an allocation, each agent has one triple from $X(i)$ or consumes his outside option, $\varnothing$. One only cares about one's own triple, so i has a preference relation over $X(i) \cup \{\varnothing\}$. We also assume that this preference is *strict*—that is, a linear order over $X(i) \cup \{\varnothing\}$—as is typical in the much of this literature.[15]

This object allocation model can be embedded into our general model as follows. An *allocation* is a subset μ of X such that no two triples name the same agent and each object's assignment is a feasible set for it. If $\mu(i)$ is empty for agent i, he consumes his outside option, $\varnothing$. That is, $\mathcal{F}$ is a subset of 2^X such that for each $\mu \in \mathcal{F}$, each $i \in N$, and each $o \in O$, $|\mu(i)| \leq 1$ and $\mu(o) \in \mathcal{F}_o$. Thus, the participants at μ, $N(\mu)$, are the agents associated with some triple in μ. Each agent's preference relation in the object allocation model defines a preference relation over $\mathcal{F}$.

Since preferences are strict, let $\boldsymbol{\mathcal{P}} \equiv \times_{i \in N} \mathcal{P}_i$, where $\boldsymbol{\mathcal{P}_i} \subseteq \mathcal{R}_i$ is the set of preferences of i over $\mathcal{F}_i$ that correspond to the strict preferences over $X(i) \cup \{\varnothing\}$. For each $R_i \in \mathcal{P}_i$, I_i is trivial and P_i completely identifies R_i. So we refer to $P_i \in \mathcal{P}_i$. Notice that $\mathcal{P}$ necessarily satisfies no indifference with the outside option. The richness assumption is much weaker than requiring $\mathcal{P}_i$ to contain all strict preferences, which is a standard assumption in such contexts.

When T is a singleton, each $x \in X$ is fully identified by the associated agent and object. In such cases, for each $i \in N$, each triple in $X(i)$ is identified by an element of O, while for each $o \in O$, each triple in $X(o)$ is identified by an element of N. Also, each object's feasible set is identified by a collection of subsets of N, while each agent's preference relation is identified by an ordering of $O \cup \{\varnothing\}$.

As an example, consider the *classical* object allocation model, where each $o \in O$ is an object with capacity $q_o \in \mathbb{Z}_+$. In this model, there is only one term under which an agent can be assigned to an object. For each $o \in O$, $\mathcal{F}_o$ consists of all subsets of $X(o)$ containing no more than q_o elements. In such cases, where for each $o \in O$, there is $q_o \in \mathbb{R}_+$ such that, for each $Y \subseteq X(o)$, $Y \in \mathcal{F}_o$ if and only if $|Y| \leq q_o$, we say that $\mathcal{F}_o$ is **capacity-based**.

This model accommodates economies with cross-object constraints: $\mathcal{F}$ can be any subset of 2^X so long as each allocation contains at most one triple associated with each agent. In the absence of such cross-object constraints, the only constraints are that the triples chosen for each object are feasible. That is, for each $\mu \subseteq X$, $\mu \in \mathcal{F}$ if and only if (*1*) for each $i \in N$, $|\mu(i)| \leq 1$, and (*2*) for each $o \in O$, $\mu(o) \in \mathcal{F}_o$. In such cases, we say that $\mathcal{F}$ is **Cartesian**. If, in addition to $\mathcal{F}$ being Cartesian, for each $o \in O$, $\mathcal{F}_o$ is capacity-based, then we say that $\mathcal{F}$ is **capacity-based.** Since we do not require $\mathcal{F}$ to be Cartesian, our results are applicable even with distributional constraints [Kamada and Kojima, 2015, Goto et al., 2017].

[15] See Bogomolnaia et al. [2005] and Erdil and Ergin [2017] for difficulties that arise with indifferences.

Non-wastefulness For the classical object allocation model, where feasibility is capacity-based, a natural requirement is that an agent ought not to prefer an object that has remaining capacity to his assignment. If he were to, we could allow him to consume this available resource at no expense to the other agents.

We define non-wastefulness for our general object allocation model as follows: Given $P \in \mathcal{P}$, $\mu \in \mathcal{F}$ is *wasteful at P* if there are $o \in O$, $i \in N$, and $\nu \in \mathcal{F}$, such that (*1*) $|\nu(o)| > |\mu(o)|$, so that ν allocates o to more agents than μ does, (*2*) $\nu(i)\ P_i\ \mu(i)$, so that i prefers his assignment at ν to that at μ, and (*3*) for each $j \in N \setminus \{i\}$, $\nu(j)\ R_i\ \mu(j)$, so that no agent is worse off at ν compared to μ. If it is not wasteful at P, then μ is *non-wasteful at P*. A mechanism, φ, is *non-wasteful* if, for each $P \in \mathcal{P}$, $\varphi(P)$ is non-wasteful at P.

Remark 3. *If an allocation is non-wasteful, then it is Pareto-constrained participation-maximal. The converse is not true, even for the classical object allocation model.*[16]

For the classical object allocation model, Balinski and Sönmez [1999] define non-wastefulness as follows: $\mu \in \mathcal{F}$ is non-wasteful at $P \in \mathcal{P}$ if there is no $o \in O$ such that $|\mu(o)| < q_o$ and $i \in N$ such that $o\ P_i\ \mu(i)$. For this narrower setting, our definition of non-wastefulness is equivalent.[17,18]

Choice In many applications, there is more information available about each object than just the feasible sets. These might be priorities over agents as in school choice, objectives of the army in cadet-branching, and so on. We model the extra information about how feasible sets are prioritized by associating each $o \in O$ with a **choice correspondence**, $\boldsymbol{C_o : 2^{X(o)} \rightrightarrows 2^{X(o)}}$, such that (*1*) for each $Y \subseteq X(o), C_o(Y) \subseteq 2^Y$, and (*2*) the range of (C_o) is mcF_o.[19] Condition (*1*) says that from any set, C_o picks only subsets of it. Condition (*2*) says that the feasible sets are exactly those that are chosen from *some* set. To satisfy Condition (*2*), it would suffice, for instance, to select each feasible set from itself.[20] Let $C \equiv (C_o)_{o \in O}$.

We associate each object with a choice correspondence rather than a *choice function*, since applications like school choice with weak priorities [Erdil and Ergin, 2008, Abdulkadiroğlu et al., 2009] are better modeled with choice correspondences.[21]

[16] See the Online Appendix for a proof.

[17] See the Online Appendix for a proof.

[18] For capacity-based $\mathcal{F}$ and singleton T, Ehlers and Klaus [2014] define an even weaker property that they call *weak non-wastefulness*. However, even in that specific setting, a result like the Structure Lemma does not hold for such a weak version of non-wastefulness.

[19] The range of $C_o : 2^{X(o)} \rightrightarrows 2^{X(o)}$ is $\bigcup_{Y \subseteq X(o)} C_o(Y)$.

[20] An alternative approach is to start with C_o as the primitive and define $\mathcal{F}_o$ to be its range.

[21] To our knowledge, the first and only other analysis of matching to consider general choice correspondences as a primitive is by Erdil and Kumano [2014].

Stability An allocation is stable if no set of agents prefers to drop their assignments in favor of being assigned to a new object under some terms that the object would "choose." That is, for each $\mu \in \mathcal{F}$ and $P \in \mathcal{P}$, μ is stable at P if it is individually rational[22] at P and there are no $o \in O$ and $Y \subseteq X(o) \setminus \mu(o)$ such that (1) for each $i \in N, |Y(i)| \leq 1$, (2) for each $y \in Y, y\ P_{N(y)}\ \mu(N(y))$, (3) $\mu(o) \notin C_o(\mu(o) \cup Y)$, (4) there is $Z \in C_o(\mu(o) \cup Y)$ such that $Y \subseteq Z$, and (5) $(\mu \setminus (\mu(o) \cup \mu(N(Y)))) \cup Z \in \mathcal{F}$. Condition (1) says that Y contains at most one triple per agent. Condition (2) says that every agent associated with a triple in Y finds it preferable to his triple in μ. These are familiar conditions from the definition of stability for choice functions. Since we are concerned with choice correspondences, the next part of the definition needs to be broken into two parts. The first, Condition (3), says that $\mu(o)$ is not among what is chosen by o when Y is available. The second, Condition (4), says that there is some chosen set, Z, that contains Y. That is, Condition (3) and Condition (4) together say that Y is contained in some Z that is *revealed* by C_o to have a higher priority than $\mu(o)$. The standard definition of stability typically does not include Condition (3) since it is implied by Condition (4) when choice correspondences are single-valued. Condition (5) is only relevant if $\mathcal{F}$ is not Cartesian. It requires that replacing the assignments of o and $N(Y)$ at μ with Z is feasible. A mechanism, φ, is *stable* if, for each $P \in \mathcal{P}$, $\varphi(P)$ is stable at P.[23]

Given a profile of preferences P, if a stable allocation μ Pareto-improves every other stable allocation at P, then it is the **agent-optimal stable** allocation at P. If a stable allocation μ is Pareto-improved by every other stable allocation at P, then it is the **agent-pessimal stable** allocation at P. If C is such that an agent-optimal stable allocation exists for each profile of preferences, then we denote by $\boldsymbol{\varphi^{AOS}}$ the mechanism that selects this allocation. Similarly, let $\boldsymbol{\varphi^{APS}}$ select the agent-pessimal stable allocation at each profile of preferences.

Stability is relevant if the choice correspondences represent more than feasibility constraints: they may represent the rights of agents with regards to the objects or particular design goals of the policy maker. The constraints imposed by this information may keep a normative benchmark rule below the Pareto frontier. Stability is a natural requirement for such a benchmark that the mechanism designer may need to Pareto-improve under true preferences. The designer's choice of mechanism might then be constrained to those that select, at every preference profile, an allocation that Pareto-improves *some* stable allocation. Such a mechanism is **stable-dominating**.[24] Since we do not insist on strict

[22] Individual rationality accounts for agents' preferences while feasibility, along with the requirement that, for each $o \in O$, $\mathcal{F}_o$ be the range of C_o, accounts for objects' choice correspondences.

[23] This definition is equivalent to the standard one if, for each $o \in O$, C_o is single-valued.

[24] We define this property for allocations as well: $\mu \in \mathcal{F}$ is *stable-dominating* if there is $\nu \in \mathcal{F}$ such that ν

Pareto-improvement, every stable mechanism is stable-dominating.

Results Since Corollary 3 only applies to Pareto-constrained participation-maximal benchmarks, we need stability to imply Pareto-constrained participation-maximality to invoke it here. In many applications, like school choice, where priority rankings define the choice correspondences, it is easy to see that stability implies non-wastefulness, and so, by Remark 3, Pareto-constrained participation-maximality. However, this may not be the case without any restrictions on choice correspondences, even if they are single-valued.[25]

We place two restrictions on choice correspondences to address this. The first says that the choices from each set should be at least as large as each choice from each of its subsets. That is, C is **size monotonic** if, for each $o \in O$, each $Y \subseteq X(o)$, each finite $Y' \subseteq Y$, each $Z \in C_o(Y)$, and each $Z' \in C_o(Y')$, $|Z| \geq |Z'|$.[26,27] The second restriction is a mild consistency requirement. It says that if a set is among those chosen from a larger set, it ought to be among what is chosen from itself. That is, C is **idempotent** if, for each $o \in O$, and each $Y \in \text{range}(C_o)$, $Y \in C_o(Y)$.[28] Unlike most of the literature on matching with contracts, we do not assume that choice correspondences satisfy a condition like *irrelevance of rejected contracts* (IRC) [Aygün and Sönmez, 2013].[29]

The assumptions that preferences are strict and that choice correspondences are size monotonic and idempotent ensure that every stable allocation is non-wasteful. The non-triviality of the proof, in the Appendix, is because we have not assumed IRC.

Lemma 2. *Let C be size monotonic and idempotent. For each profile of preferences, if an allocation is stable, then it is non-wasteful.*

Lemma 2 allows us to link our results to the literature on stable mechanisms. In particular, we have the following corollary of Theorem 1, Lemma 2, and Remark 3.

Corollary 4. *Let C be size monotonic and idempotent. For each stable-dominating benchmark rule, there is at most one strategy-proof mechanism that Pareto-improves it under true preferences.*

is stable and μ Pareto-improves ν.

[25] See the Online Appendix for an example.

[26] This is an extension to correspondences of a condition defined for choice functions [Alkan, 2002, Alkan and Gale, 2003, Fleiner, 2003, Hatfield and Milgrom, 2005].

[27] For finite Y, setting $Y' = Y$, size monotonicity implies that for each pair $Z, Z' \in C_o(Y)$, $|Z| = |Z'|$.

[28] This rules out, for instance, C_o such that $C_o(\{x,y,z\}) = \{\{x,y\}\}$ but $C_o(\{x,y\}) = \{\{x\}\}$.

[29] Aygün and Sönmez [2013] define IRC for choice functions. Alva [2018] extends the definition of IRC to choice correspondences and shows its equivalence to the weak axiom of revealed preference.

Discussion An immediate implication of Corollary 4 is that every strategy-proof and stable-dominating mechanism is strategy-proofness-constrained Pareto-efficient. Hirata and Kasuya [2017] define a version of non-wastefulness weaker than ours, but is logically independent from Pareto-constrained participation-maximality. They show that for matching with contracts models with single-valued choice functions satisfying IRC, every non-wasteful and strategy-proof mechanism is strategy-proofness-constrained Pareto-efficient. They also show that for such choice functions, there is at most one stable and strategy-proof mechanism. Under our choice assumptions, there may be more than one such mechanism.[30]

Kamada and Kojima [2015], in a model with distributional constraints, define a strategy-proof mechanism that Pareto-improves on a deferred acceptance mechanism. This does not contradict Corollary 4, which takes $\mathcal{F}$ as fixed. By making flexible the constraints that define $\mathcal{F}$, they obtain a strategy-proof Pareto-improvement on a benchmark that is no longer Pareto-constrained participation-maximal under the redefined $\mathcal{F}$.[31] In a similar setting with more general constraints, Goto et al. [2017] show that an adaptation of the deferred acceptance mechanism is strategy-proofness-constrained Pareto-efficient.

A simple yet policy-relevant measure of a mechanism's performance is the *expected match rate*: the expected number, with respect to a prior distribution over preference profiles, of agents to whom the mechanism assigns an object. Theorem 1 implies that comparisons of strategy-proof mechanisms by Pareto-improvement translate to comparisons of expected match rates. An interesting application concerns the proposal of Kamada and Kojima [2015] to address doctor shortages in rural areas [Roth, 1986]. Their strategy-proof mechanism assigns more doctors to rural areas if only caps on the number of doctors assigned to non-rural regions are tightened. However, this tightening leads to a Pareto-worse outcome from the doctors' point of view. Thus, Theorem 1 provides a cautionary message: an expected increase of one doctor matched to a rural region leads to an expected decrease of more than one doctor matched to other regions.[32]

Many of the mechanisms that solve market design problems are based on the cumulative offer algorithm [Hatfield and Milgrom, 2005]. Under size-monotonic and idempotent

[30] See the Online Appendix for an example.

[31] In particular, the benchmark rule in their study is the deferred acceptance mechanism where objects have target capacities that satisfy distributional constraints. Taking these targets as given, this benchmark has no strategy-proof Pareto-improvement. However, by allowing a mechanism the flexibility to exceed these targets while satisfying the distributional constraints, which changes the set of feasible allocations $\mathcal{F}$, Kamada and Kojima [2015] obtain a strategy-proof Pareto-improvement.

[32] The conclusion that the decrease is greater than one holds if the prior distribution over preference profiles has full support.

choice functions,[33] these mechanisms are strategy-proofness-constrained Pareto-efficient, by Corollary 4. However, these choice conditions are not necessary. To apply Corollaries 2 or 3, one only needs to ensure non-wastefulness or Pareto-constrained participation-maximality. Typically, it is easy to show that these mechanisms are non-wasteful even without size-monotonicity.[34] Thus, they are also strategy-proofness-constrained Pareto-efficient.

5.2 School Choice

We consider here the more specialized *school choice* model, which is the classical object allocation model augmented with weak priorities. In this model, T is a singleton and $\mathcal{F}$ is capacity-based. Additionally, each $o \in O$ is associated with a **priority** over N denoted by $\boldsymbol{\succsim_o}$, which is a complete, transitive, and reflexive binary relation. Let $\boldsymbol{\succsim} \equiv (\succsim_o)_{o \in O}$.

Given priorities, $\succsim$, for each $o \in O$, we define $C_o^{\succsim}$ as follows. For each $Y \subseteq N$,

$$C_o^{\succsim}(Y) \equiv \begin{cases} \{Y\} & \text{if } |Y| \leq q_o, \\ \{Z \subseteq Y : |Z| = q_o \text{ and for each } i \in Z \text{ and each } j \in Y \setminus Z, i \succsim_o j\} & \text{otherwise.} \end{cases}$$

That is, for each subset of agents, if it contains no more than q_o elements, then the entire set is the only one that is chosen. If not, then all subsets that contain exactly q_o elements are chosen, except for ones that include agents with *strictly* lower priority than an excluded agent. This $C_o^{\succsim}$ is size monotonic and idempotent.

Suppose that an agent prefers a particular object o to the one that he is assigned. Under the interpretation of priorities as consumption "rights," if o is assigned to someone else who has *strictly* lower priority, then i has the right to protest this allocation. For each $\mu \in \mathcal{F}$, μ **respects priorities** if no agent can protest on such grounds. That is, there is no pair $i, j \in N$ and $o \in O$ such that $o\ P_j\ \mu(j)$, $\mu(i) = o$, and $j \succ_o i$.

Stability and dominating stable allocations as fairness Interpreting respect for the priorities as a fairness constraint [Balinski and Sönmez, 1999, Abdulkadiroğlu and Sönmez, 2003], we are interested in mechanisms that are individually rational, non-wasteful, and fair. Respect for priorities alongside individual rationality and non-wastefulness is equivalent to stability with respect the choice correspondence for each object defined from these priorities.

[33] See, for instance, the problems of cadet-branch matching [Sönmez and Switzer, 2013, Sönmez, 2013] and controlled school choice [Hafalir et al., 2013, Ehlers et al., 2014].

[34] These include problems with slot-specific priorities [Kominers and Sönmez, 2016] or hidden substitutes [Hatfield and Kominers, 2014].

Remark 4. *For each profile of preferences, an allocation is stable with respect to $C^{\succsim}$ if and only if it is individually rational, non-wasteful, and respects $\succsim$.*[35]

Since the combination of individual rationality, non-wastefulness, and respect for priorities is equivalent to stability, we speak of an allocation being stable rather than it respecting priorities and being individually rational and non-wasteful.

When priorities are strict, φ^{AOS} is well defined [Gale and Shapley, 1962]. However, when priorities contain ties, there may not exist a single stable allocation that Pareto-improves every other stable allocation. Since φ^{AOS} may not be well defined, a common approach to handling weak priorities is to use tie breakers to form strict priorities. Let $\tau \equiv (\tau_o)_{o \in O}$ be a profile of linear orders over N, one for each object. For each such τ, let $\succsim^{\boldsymbol{\tau}}$ be the priorities **tie broken by $\boldsymbol{\tau}$**. That is, for each distinct pair $i, j \in N$, $i \succ_o^{\tau} j$ if either (1) $i \succ_o j$ or (2) $i \sim_o j$ and $i \; \tau_o \; j$. Let $\boldsymbol{\mathcal{T}}$ be the set of all profiles of tie breakers. Since the agent-optimal stable mechanism for strict priorities is well defined, given $\succsim$ and $\tau \in \mathcal{T}$, we define the agent-optimal stable mechanism for the priorities tie broken by τ as $\boldsymbol{\varphi^{AOS\tau}}$. These are the mechanisms studied by Abdulkadiroğlu et al. [2009].[36]

For strict priorities, unless they satisfy a restrictive acyclicity condition, φ^{AOS} is not Pareto-efficient [Ergin, 2002]. So when priorities are weak, arbitrarily breaking ties could cause Pareto-inefficiency. In fact, for any $\tau \in \mathcal{T}$, $\varphi^{AOS\tau}$ may select an allocation that is Pareto-improvable by another stable allocation [Erdil and Ergin, 2008]. Furthermore, some priorities permit a stable, (group) strategy-proof, and Pareto-efficient mechanism, even while $\varphi^{AOS\tau}$ is Pareto-inefficient for every $\tau \in \mathcal{T}$ [Ehlers and Erdil, 2010].

Since $\varphi^{AOS\tau}$ may not be Pareto-efficient, Abdulkadiroğlu et al. [2009] consider the following question: for each $\tau \in \mathcal{T}$, is it possible to find a strategy-proof mechanism that Pareto-improves $\varphi^{AOS\tau}$? They show that the answer is negative.[37] However, even if the answer were positive, the allocations chosen by the Pareto-improving mechanism would not be stable. The reason such a mechanism would pass muster here is that it Pareto-improves a *stable* mechanism: $\varphi^{AOS\tau}$. But as we have explained above, there is nothing special about $\varphi^{AOS\tau}$, other than strategy-proofness, when priorities are weak. In fact, they may even select allocations that are Pareto-improved by other stable allocations.

In Alva and Manjunath [2017], we suggest the choice of a mechanism can be justified on the grounds that it Pareto-improves *some* stable allocation at each profile of preferences. If an agent were to protest the violation of his priority at some object, we offer a

[35] See the Online Appendix for a proof.

[36] These $\varphi^{AOS\tau}$s correspond to what Abdulkadiroğlu et al. [2009] call deferred acceptance with multiple tie-breaking.

[37] Kesten and Kurino [2015] shows the same result even without outside options.

move to the Pareto-improved stable allocation so this protest would be moot: the agent would not better off at *this* stable allocation.[38] That is, by requiring the chosen mechanism to just be *stable-dominating*, rather than stable, we may enlarge the options for a strategy-proof Pareto-improvement.

For strict priorities, since φ^{APS} is well defined [Gale and Shapley, 1962], Corollary 4 implies the following, which is stronger than the known result that φ^{AOS} is the only stable and strategy-proof mechanism [Alcalde and Barberà, 1994].

Corollary 5. *If $\succsim$ consists of strict priorities, then φ^{AOS} is the unique stable-dominating and strategy-proof mechanism.*

On the other hand, for weak priorities, there may be more than one stable-dominating and strategy-proof mechanism. Nevertheless, Proposition 1 yields the following corollary.

Corollary 6. *Each stable-dominating and strategy-proof mechanism (including, for each $\tau \in \mathcal{T}$, $\varphi^{AOS\tau}$) is strategy-proofness-constrained Pareto-efficient. Furthermore, no two stable-dominating and strategy-proof mechanisms are IR-PCPM-Pareto-connected.*

While, for each $\tau \in \mathcal{T}$, $\varphi^{AOS\tau}$ is strategy-proofness-constrained Pareto-efficient, $\varphi^{AOS\tau}$s are not the only stable and strategy-proof mechanisms [Ehlers and Erdil, 2010]. Corollary 6 extends the result of Abdulkadiroğlu et al. [2009] to all of these.

Beyond stability as fairness While we have focused on stability and stable-domination as notions of fairness, our results allow us to draw conclusions about other fairness concepts as well. Take for instance the *legal set* of allocations under strict priorities [Morrill, 2016], where only *harmful* and *redressable* priority violations are ruled out.[39] The legal set includes the stable set and always has a Pareto-worst member that is non-wasteful. Consequently, Proposition 1 implies that φ^{AOS} is the unique strategy-proof selection from the legal set.

Corollary 7. *If $\succsim$ consists solely of strict priorities, then φ^{AOS} is the unique strategy-proof mechanism selecting from the legal set.*

[38] See Dur et al. [2015] for recent work on algorithms that select from the stable-dominating set in a way that only certain priorities are violated.

[39] We refer the reader to Morrill [2016] for a formal definition.

5.3 Excludable Public Goods

In this section, we illustrate how to accommodate excludable public goods in our model. We focus on a simple setting for expositional clarity, since our purpose is to highlight how to apply our results to obtain novel insights. A thorough analysis of the efficient frontier of strategy-proof mechanisms for a general excludable public goods model could use the full power of our results, but would take us too far afield.

The following model is similar to that of Jackson and Nicolò [2004] and Cantala [2004]. Suppose a public facility is to be located on the interval $(0,1)$ and a set of users chosen. The set of agents is partitioned into two sets: agents in $\boldsymbol{N_L}$ live at 0 and agents in $\boldsymbol{N_R}$ live at 1. Each $i \in N$ prefers to have the facility located as close to his own home as possible. Furthermore, he is unwilling to travel beyond a certain **threshold** $\boldsymbol{t_i} \in (0,1)$. That is, if i lives at 0, then he is unwilling to travel to the right of t_i and if he lives at 1, then he is unwilling to travel to the left of t_i. Since we know exactly how each agent ranks the locations, the only private information for i is t_i. So a preference profile is identified by the threshold profile t. We assume that each $i \in N$ prefers to travel to t_i and enjoy the facility over opting out. However, he prefers to opt out rather than travel beyond t_i. Thus, no indifference with the outside option is satisfied. Since $t_i \in (0,1)$, there is always some positive distance he is willing to travel. Yet, richness of the outside option is satisfied since t_i may be arbitrarily close to i's home (0 if $i \in N_L$ or 1 if $i \in N_R$).

An *allocation* consists of two parts: the location of a public facility in the interval $(0,1)$ and the set of users. That is, $\mathcal{F} \equiv [0,1] \times 2^N$. A mechanism maps threshold profiles to allocations. It is **dictatorial** if it ignores the thresholds and always locates the facility at 0 or always locates the facility at 1.

Even in this stark model, if we insist on strategy-proofness, individually rationality, and Pareto-efficiency, then the only two options are the dictatorial mechanisms.[40] Thus, the requirement of Pareto-efficiency alongside strategy-proofness and individual rationality precludes the possibility that any compromise is ever reached.

Are there attractive mechanisms that compromise if we give up Pareto-efficiency? Consider the following family of mechanisms that select a compromise location at some threshold profiles. A *one-sided unanimous compromise* mechanism is defined by a compromise $x \in (0,1)$ and a side $J \in \{L, R\}$. If agents in J unanimously find x acceptable, then the mechanism selects the location x. Otherwise, it selects the home location of the other side, denoted K. The set of users is the set of agents willing to travel to the chosen location.

Each one-sided unanimous compromise mechanism is individually rational by defi-

[40] See the Online Appendix for a proof.

nition. It is also strategy-proof: only threshold reports of J are used to determine the location, and it is clear that misreports can only hurt an agent in J, independently of others' reports. However, it is not Pareto-efficient: if every agent in K finds compromise x unacceptable yet every agent in J finds it acceptable, then x is chosen by the mechanism even though the home location of agents in J is a Pareto-improvement.

When the chosen location is x, every agent in N_J participates. An agent in N_K who does not participate would do so only if the location were moved some distance towards his home location. But this would make every member of N_J worse off. On the other hand, if the chosen location is the home location of K, then any other location would make members of N_K worse off. Therefore, every one-sided compromise mechanism is Pareto-constrained participation-maximal. So, by Corollary 3, it is on the Pareto-frontier of strategy-proof mechanisms.

Corollary 8. *Each one-sided unanimous compromise mechanism is strategy-proofness-constrained Pareto-efficient.*

5.4 Transferable Utility

Here we consider the problem of making a social decision and assigning payments to agents when preferences are quasilinear in payments. Consequently, there is indifference with the outside option. Nonetheless, Corollary 1 applies if richness of the outside option is satisfied. Below we study its implications.

Suppose $\boldsymbol{\mathcal{D}}$ is a set of **social decisions**. At each $\delta \in \mathcal{D}$, the agents participating in δ are $\boldsymbol{N(\delta)}$. Let $\boldsymbol{\mathcal{T}} \subseteq \mathbb{R}^N$ be the set of **possible payment profiles**. At each $\tau \in \mathcal{T}$, for each $i \in N$, τ_i is the payment that i makes. Like $\mathcal{F}_i$, let $\boldsymbol{\mathcal{D}_i}$ be the set of decisions that i participates at.

To fit this into our general model, an allocation in $\mathcal{F}$ is a pair $(\delta, \tau) \in \mathcal{D} \times \mathcal{T}$. Furthermore, $N(\delta, \tau) = N(\delta)$. Since a non-participant consumes his outside option, we require that for each $i \notin N(\delta)$, $\tau_i = 0$. That is, no transfers are made to or from non-participants.

We assume that each agent's preferences are quasilinear in the payment that he makes. Thus, for each $i \in N$, each $R_i \in \mathcal{R}_i$ is identified by a **valuation**, $\boldsymbol{v_i} \in \mathbb{R}^{\mathcal{D}_i \cup \{\varnothing\}}$. Let i's **valuation space**, $\boldsymbol{\mathcal{V}_i}$, be the set of all possible valuations for i. Thus, i's preference relation is represented by $v_i(\delta) - \tau_i$, where $v_i(\delta)$ is the δth coordinate of v_i if $i \in N(\delta)$ and the $\varnothing$th coordinate otherwise. The set of all **valuation profiles** is $\boldsymbol{\mathcal{V}} \equiv \times_{i \in N} \mathcal{V}_i$. For each $v \in \mathcal{V}$, $i \in N$, and $\delta \in \mathcal{D}$, denote by $\boldsymbol{u_i^v(\delta)}$ the **net value** of δ to i, relative to being excluded. That is, $u_i^v(\delta) \equiv v_i(\delta) - v_i(\varnothing)$. Interpreting $v_i(\varnothing)$ as i's opportunity cost of participating, $u_i^v(\delta)$ is i's valuation of δ net of this cost. Of course, if $i \notin N(\delta)$, then $u_i^v(\delta) = 0$.

The richness of the outside option condition that we described in Section 2 was for preference relations over $\mathcal{F}_i \cup \{\varnothing\}$. It would be met if, for instance, $\mathcal{V}_i$ were such that for each $v_i \in \mathcal{V}_i$ there were $v_i' \in \mathcal{V}_i$, such that the valuations of each of the decisions were unchanged, but the valuation of the outside option were increased. That is, for each $\kappa > v_i(\varnothing)$, there is v_i' such that $v_i'(\varnothing) = \kappa$ and for each $\delta \in \mathcal{D}_i$, $v_i'(\delta) = v_i(\delta)$.[41]

A mechanism in this setting consists of two parts: a **decision rule**, $d : \mathcal{V} \to \mathcal{D}$, and a **payment rule**, $t : \mathcal{V} \to \mathcal{T}$. If (d,t) is strategy-proof, we say that t **implements** d. If there is a payment rule that implements d, then d is **implementable**. As defined, implementation says nothing about individual rationality. We define parallel concepts with this requirement added: t **IR-implements** d if (d,t) is not only strategy-proof but also individually rational. In this case, we say that d is **IR-implementable**. If there is a unique payment rule that IR-implements d, we say that d is **uniquely IR-implementable**.

An **efficient** decision rule maximizes the net value of the decision to its participants. That is, a decision rule d is efficient if, for each $v \in \mathcal{V}$,

$$d(v) \in \operatorname*{argmax}_{\delta \in \mathcal{D}} \sum_{i \in N} u_i^v(\delta).$$

An efficient mechanism is one that has an efficient decision rule. In a context where the status quo is that each agent enjoys his outside option, an efficient decision maximizes the total surplus over the status quo.

With these concepts in hand, we obtain the following counterpart to Corollary 1:

Corollary 9. *Every IR-implementable decision rule is uniquely IR-implementable.*

In fact we can say more than Corollary 9. For any pair of decision rules d and d' that are participation-equivalent, if t IR-implements d and t' IR-implements d', then (d,t) and (d',t') are welfare-equivalent.

This leads to a characterization of **pivotal mechanisms**: efficient mechanisms with *pivotal* payment rules. A pivotal payment rule assigns to each agent a payment equal to the externality that his participation imposes on others: t is a pivotal payment rule if, for each $v \in \mathcal{V}$ and each $i \in N$,

$$t_i(v) \equiv \max_{\delta \in \mathcal{D} \setminus \mathcal{D}_i} \left\{ \sum_{j \neq i} u_j^v(\delta) \right\} - \sum_{j \neq i} u_j^v(d(v)).$$

A pivotal payment rule is a particular Groves scheme [Groves, 1973] and pivotal

[41] We could state this with a bound on κ if there were a bound on valuations of the decisions in $\mathcal{D}_i$ and on the payments that i may make in $\mathcal{T}$.

mechanisms are well understood to be strategy-proof [Vickrey, 1961, Clarke, 1971]. Moreover, for each $v \in \mathcal{V}$ and each $i \in N$, if $i \notin N(d(v))$, then t prescribes a zero payment for i. Therefore, (d,t) is feasible. Finally, the efficiency of d ensures that (d,t) is also individually rational. Thus, from Corollary 9 we have:

Corollary 10. *A mechanism (d,t) is efficient, strategy-proof, and individually rational if and only if it is a pivotal mechanism. That is, t IR-implements efficient d if and only if it is the pivotal payment rule.*

When we have neither the restriction that non-participants receive zero payments nor the requirement of individual rationality, Groves schemes are the only ones to implement efficient decision rules [Green and Laffont, 1977, Holmström, 1979]. For private goods economies, pivotal rules are the only Groves schemes that are individually rational without making payments to non-participants [Chew and Serizawa, 2007]. For pure public goods, the counterpart of individual rationality requires that agents have no incentive to free-ride. Substituting this property for individual rationality similarly characterizes pivotal rules [Moulin, 1986]. Corollary 10 neither implies nor is implied by existing characterizations of the pivotal mechanism since our assumption that $\mathcal{V}$ satisfy richness of the outside option is logically independent from the domain assumptions made in existing results.

Appendix: Proof of Lemma 2

Before showing that stability implies non-wastefulness under the assumptions of size monotonicity and idempotence, we start with a definition and a lemma. For each $P \in \mathcal{P}$, each $\mu \in \mathcal{F}$, and each $o \in O$, let $\boldsymbol{Y_o^\mu(P)}$ be the triples in $X(o)$ that are associated with an agent who prefers it to what he is assigned at μ. That is, $Y_o^\mu(P) \equiv \{x \in X(o) : x \; P_{N(x)} \; \mu(N(x))\}$.

Lemma 3. *Let C be size monotonic and idempotent. For each $P \in \mathcal{P}$, each stable $\mu \in \mathcal{F}$, each $o \in O$, each finite $Y \subseteq Y_o^\mu(P)$, and each $Z \in C_o(\mu(o) \cup Y)$, $|Z| = |\mu(o)|$.*

Proof. We proceed by induction over subsets of $Y_o^\mu(P)$. Let $Y \subseteq Y_o^\mu(P)$.

For the base case, where $Y = \varnothing$, since $\mu(o) \in \text{range}(C_o)$, by idempotence of C, $\mu(o) \in C_0(\mu(o) \cup Y)$ and by size monotonicity of C, for each $Z \in C_o(\mu(o))$, $|Z| = |\mu(o)|$.

As an induction hypothesis, assume that for each $Y' \subsetneq Y$ and each $Z \in C_o(\mu(o) \cup Y')$, $|Z| = |\mu(o)|$. Equivalently, for each $T \subseteq \mu(o) \cup Y$ such that $Y \not\subseteq T$, for each $Z \in C_o(\mu(o) \cup T)$, $|Z| = |\mu(o)|$.

The induction step is to show that, for each $Z \in C_o(\mu(o) \cup Y)$, $|Z| = |\mu(o)|$. Let $Z \in C_o(\mu(o) \cup Y)$. By idempotence of C, $Z \in C_o(Z)$. Thus, since $Z \subseteq \mu(o) \cup Z$, by size monotonicity of C,

$$\text{for each } Z' \in C_o(\mu(o) \cup Z), |Z| \leq |Z'|. \tag{1}$$

By stability of μ, either $Y \not\subseteq Z$ or $\mu(o) \in C_o(\mu(o) \cup Y)$. If $\mu(o) \in C_o(\mu(o) \cup Y)$, then by size monotonicity of C, for each $Z \in C_o(\mu(o) \cup Y)$, $|Z| = |\mu(o)|$. Instead, if $Y \not\subseteq Z$, the induction hypothesis implies

$$\text{for each } Z' \in C_o(\mu(o) \cup Z), |Z'| = |\mu(o)|. \tag{2}$$

By (1) and (2), $|Z| \leq |\mu(o)|$. Since C is idempotent and $\mu \in \mathcal{F}$, $\mu(o) \in C_o(\mu(o))$. Then, since C is size monotonic and $Z \in C_o(\mu(o) \cup Y)$, $|\mu(o)| \leq |Z|$. Thus, we conclude that $|Z| = |\mu(o)|$. $\square$

Proof of Lemma 2. Suppose that μ is wasteful. Then there are $o \in O$ and $\nu \in \mathcal{F}$ such that $|\nu(o)| > |\mu(o)|$ and, for each $y \in \nu(o) \setminus \mu(o)$, $y \; P_{N(y)} \; \mu(N(y))$. Let $Y \equiv \nu(o) \setminus \mu(o)$. Since $Y \subseteq Y_o^{\mu}(P)$ and Y is finite, by Lemma 3, for each $Z' \in C_o(\mu(o) \cup Y)$, $|Z'| = |\mu(o)|$. However, $\nu(o) \subseteq \mu(o) \cup Y$. So by size monotonicity of C, for each $Z \in C_o(\nu(o))$ and each $Z' \in C_o(\mu(o) \cup Y)$, $|Z| \leq |Z'| = |\mu(o)|$. Since $\nu \in \mathcal{F}$, $\nu \in \mathcal{F}_o = \text{range}(C_o)$. So by idempotence of C, $\nu(o) \in C_o(\nu(o))$. Thus, $|\nu(o)| \leq |\mu(o)|$. This contradicts the definition of ν. $\square$

References

A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth. Strategy-proofness versus efficiency in matching with indifferences: Redesigning the NYC high school match. *American Economic Review*, 99(5):1954–1978, December 2009. [13], [18], [19]

A. Abdulkadiroğlu and T. Sönmez. School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747, June 2003. [1], [17]

J. Alcalde and S. Barberà. Top dominance and the possibility of strategy-proof stable solutions to matching problems. *Economic Theory*, 4(3):417–35, May 1994. [19]

A. Alkan. A class of multipartner matching markets with a strong lattice structure. *Economic Theory*, 19(4):737–746, June 2002. [15]

A. Alkan and D. Gale. Stable schedule matching under revealed preference. *Journal of Economic Theory*, 112(2):289–306, October 2003. [15]

S. Alva. WARP and combinatorial choice. *Journal of Economic Theory*, 173:320–333, January 2018. [15]

S. Alva and V. Manjunath. Stable-dominating rules. Working paper, University of Texas at San Antonio, 2017. [9], [18]

H. Anno and M. Kurino. On the operation of multiple matching markets. *Games and Economic Behavior*, 100:166–185, November 2016. [10]

H. Anno and H. Sasaki. Second-best efficiency of allocation rules: Strategy-proofness and single-peaked preferences with multiple commodities. *Economic Theory*, 54(3): 693–716, November 2013. [10]

O. Aygün and T. Sönmez. Matching with contracts: Comment. *American Economic Review*, 103(5):2050–51, August 2013. [15]

M. Balinski and T. Sönmez. A tale of two mechanisms: Student placement. *Journal of Economic Theory*, 84(1):73–94, January 1999. [13], [17]

A. Bogomolnaia, R. Deb, and L. Ehlers. Strategy-proof assignment on the full preference domain. *Journal of Economic Theory*, 123(2):161–186, August 2005. [12]

D. Cantala. Choosing the level of a public good when agents have an outside option. *Social Choice and Welfare*, 22(3):491–514, June 2004. [20]

S. H. Chew and S. Serizawa. Characterizing the Vickrey combinatorial auction by induction. *Economic Theory*, 33:393–406, November 2007. [23]

E. H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, September 1971. [23]

U. Dur, A. A. Gitmez, and O. Yılmaz. School choice under partial fairness. Working paper, North Carolina State University, 2015. [3], [19]

L. Ehlers and A. Erdil. Efficient assignment respecting priorities. *Journal of Economic Theory*, 145(3):1269–1282, May 2010. [18], [19]

L. Ehlers and B. Klaus. Strategy-proofness makes the difference: Deferred-acceptance with responsive priorities. *Mathematics of Operations Research*, 39(4):949–966, November 2014. [13]

L. Ehlers, I. E. Hafalir, M. B. Yenmez, and M. A. Yildirim. School choice with controlled choice constraints: Hard bounds versus soft bounds. *Journal of Economic Theory*, 153: 648 – 683, September 2014. [17]

A. Erdil. Strategy-proof stochastic assignment. *Journal of Economic Theory*, 151:146 – 162, May 2014. [2], [8]

A. Erdil and H. I. Ergin. What's the matter with tie-breaking? Improving efficiency in school choice. *American Economic Review*, 98(3):669–89, June 2008. [13], [18]

A. Erdil and H. I. Ergin. Two-sided matching with indifferences. *Journal of Economic Theory*, 171:268 – 292, September 2017. [6], [12]

A. Erdil and T. Kumano. Prioritizing diversity in school choice. Working paper, University of Cambridge, 2014. [13]

H. I. Ergin. Efficient resource allocation on the basis of priorities. *Econometrica*, 70(6): 2489–2497, 2002. [18]

T. Fleiner. A fixed-point approach to stable matchings and some applications. *Mathematics of Operations Research*, 28:103–126, February 2003. [15]

D. Gale and L. Shapley. College admissions and the stability of marriage. *American Mathematical Monthly*, 69:9–15, 1962. [18], [19]

M. Goto, F. Kojima, R. Kurata, A. Tamura, and M. Yokoo. Designing matching mechanisms under general distributional constraints. *American Economic Journal: Microeconomics*, 9(2):226–262, May 2017. [12], [16]

J. Green and J.-J. Laffont. Characterization of satisfactory mechanisms for the revelation of preferences for public goods. *Econometrica*, 45:417–438, March 1977. [23]

T. Groves. Incentives in teams. *Econometrica*, 41(4):617–631, July 1973. [22]

I. E. Hafalir, M. B. Yenmez, and M. A. Yildirim. Effective affirmative action in school choice. *Theoretical Economics*, 8(2):325–363, May 2013. [17]

J. W. Hatfield and S. D. Kominers. Hidden substitutes. Working paper, University of Texas at Austin, 2014. [17]

J. W. Hatfield and P. Milgrom. Matching with contracts. *American Economic Review*, 95 (4):913–935, September 2005. [1], [11], [15], [16]

D. Hirata and Y. Kasuya. On stable and strategy-proof rules in matching markets with contracts. *Journal of Economic Theory*, 168:27–43, March 2017. [3], [16]

B. Holmström. Groves' scheme on restricted domains. *Econometrica*, 47(5):1137–1144, September 1979. [23]

A. Hylland and R. Zeckhauser. The efficient allocation of individuals to positions. *Journal of Political Economy*, 87(2):293–314, April 1979. [1]

M. O. Jackson and A. Nicolò. The strategy-proof provision of public goods under congestion and crowding preferences. *Journal of Economic Theory*, 115(2):278–308, April 2004. [1], [20]

Y. Kamada and F. Kojima. Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, 105(1):67–99, January 2015. [12], [16]

O. Kesten and M. Kurino. Do outside options matter in matching? A new perspective on the trade-offs in school choice. Working paper, Carnegie Mellon University, 2015. [18]

S. D. Kominers and T. Sönmez. Matching with slot-specific priorities: Theory. *Theoretical Economics*, 11(2):683–710, May 2016. [17]

T. Morrill. Which school assignments are legal? Working paper, North Carolina State University, 2016. [3], [19]

H. Moulin. Characterizations of the pivotal mechanism. *Journal of Public Economics*, 31 (1):53–78, October 1986. [23]

S. Ohseto. Characterizations of strategy-proof and fair mechanisms for allocating indivisible goods. *Economic Theory*, 29(1):111–121, September 2006. [10]

A. E. Roth. On the allocation of residents to rural hospitals: A general property of two-sided matching markets. *Econometrica*, 54(2):425–427, March 1986. [9], [16]

T. Sönmez. Strategy-proofness and essentially single-valued cores. *Econometrica*, 67(3): 677–690, May 1999. [6], [7]

T. Sönmez. Bidding for army career specialties: Improving the ROTC branching mechanism. *Journal of Political Economy*, 121(1):186–219, March 2013. [17]

T. Sönmez and T. B. Switzer. Matching with (branch-of-choice) contracts at the United States Military Academy. *Econometrica*, 81(2):451–488, March 2013. [17]

Y. Sprumont. Constrained-optimal strategy-proof assignment: Beyond the Groves mechanisms. *Journal of Economic Theory*, 148(3):1102–1121, May 2013. [10]

W. Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, March 1961. [23]