DISCUSSION PAPER SERIES

DP16872

Arrovian Efficiency and Auditability in Discrete Mechanism Design

Marek Pycia and M. Utku Ünver

ORGANIZATIONAL ECONOMICS

CEPR

ISSN 0265-8003

Arrovian Efficiency and Auditability in Discrete Mechanism Design

Marek Pycia and M. Utku Ünver

Discussion Paper DP16872
Published 10 January 2022
Submitted 31 December 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Arrovian Efficiency and Auditability in Discrete Mechanism Design

Abstract

We study mechanism design and preference aggregation in environments in which the space of social alternatives is discrete and the preference domain is rich, as in standard models of social choice and so-called allocation without transfers. We show that a mechanism (or aggregation rule) selects the best outcome with respect to some resolute Arrovian social welfare function if, and only if, it is Pareto efficient and auditable. We further show that auditability implies non-bossiness and is implied by the conjunction of non-bossiness and individual strategy-proofness, and that the later conjunction is equivalent to group strategy-proofness as well as to Maskin monotonicity. As applications, we derive new characterizations in voting and allocation domains.

JEL Classification: N/A

Keywords: N/A

Marek Pycia - marek.pycia@econ.uzh.ch
University of Zurich and CEPR

M. Utku Ünver - unver@bc.edu
Boston College

Arrovian Efficiency and Auditability in Discrete Mechanism Design*

Marek Pycia[†] M. Utku Ünver[‡]

This draft: November 2021

First draft: October 2013

Abstract

We study mechanism design and preference aggregation in environments in which the space of social alternatives is discrete and the preference domain is rich, as in standard models of social choice and so-called allocation without transfers. We show that a mechanism (or aggregation rule) selects the best outcome with respect to some resolute Arrovian social welfare function if, and only if, it is Pareto efficient and auditable. We further show that auditability implies non-bossiness and is implied by the conjunction of non-bossiness and individual strategy-proofness, and that the later conjunction is equivalent to group strategy-proofness as well as to Maskin monotonicity. As applications, we derive new characterizations in voting and allocation domains.

Keywords: Strategy-proofness, Pareto efficiency, Arrovian preference aggregation, auditability, non-bossiness, voting, house allocation.

JEL classification: C78, D78

*This paper subsumes the earlier preliminary results and analysis on Arrovian efficiency first reported in the October 2013 draft of Pycia and Ünver's "Incentive Compatible Allocation and Exchange of Discrete Resources." Between 2016 and 2019, an earlier version of this paper was posted under the title "Arrovian Efficiency in Allocation of Discrete Resources." We would like to thank Vikram Manjunath and audiences in Princeton and Montreal for their comments. Pycia gratefully acknowledges the support of UCLA, where he was a faculty member when working on this paper, as well as the financial support from the William S. Dietrich II Economic Theory Center at Princeton University.

†University of Zurich, Department of Economics, e-mail: marek.pycia@econ.uzh.ch.

‡Boston College, Department of Economics, e-mail: unver@bc.edu.

1 Introduction

Microeconomic theory has informed the design of many markets and other institutions. Many new mechanisms have been proposed to allocate resources in environments in which transfers are not used or are prohibited. These environments include the allocation and exchange of transplant organs, such as kidneys (Roth, Sönmez and Ünver, 2004); the allocation of school seats in Boston, New York City, Chicago, etc. (Abdulkadiroğlu and Sönmez, 2003); and the allocation of dormitory rooms at US colleges (Abdulkadiroğlu and Sönmez, 1999). The mechanisms used elicit ordinal preferences of participants.[1]

The central concerns in the development of allocation mechanisms are incentives and efficiency.[2] The literature focused on Pareto efficiency: a social alternative is Pareto efficient if there exists no other social alternative that makes everybody weakly better off and at least one individual better off.[3] Pareto efficiency however is a weak efficiency concept; while interpersonal utility comparisons are not needed for Pareto efficiency, it only gives a lower bound for what can be achieved through desirable mechanisms. In consequence, welfare economics—starting with Bergson (1938), Samuelson (1947), and Arrow (1963)—have long looked at stronger efficiency concepts requiring an efficient outcome to be the maximum of a social ranking of outcomes; an idea later named as resoluteness.[4] For instance, Arrow (1963), pp. 36-37, discusses the partial ordering of outcomes given by Pareto dominance, and observes:

> But though the study of maximal alternatives is possibly a useful preliminary to the analysis of particular social welfare functions, it is hard to see how any policy recommendations can be based merely on a knowledge of maximal alternatives. There is no way of deciding which maximal alternative to decide on.

Our paper carries out the Bergson-Samuelson-Arrow's program of analyzing stronger welfare criteria to discrete mechanism design, in which continuous transfers are not allowed and there is a finite number of alternatives. We study a broad class of discrete environments, merely imposing a natural richness assumption on preference domains; richness is a substantially weakening of Arrovian universal domain assumption and it is satisfied in many practically and theoretically relevant economic domains such as voting for candidates or issues with universal strict prefer-

[1] In the context of deterministic mechanisms without transfers eliciting ordinal information is all we can do. In addition, eliciting ordinal preferences is considered simpler and more practical (see Bogomolnaia and Moulin, 2001).

[2] For instance, Bogolomania and Moulin (2004) write that "the central question of that literature is to characterize the set of efficient and incentive compatible (strategy-proof) assignment mechanisms."

[3] Relatedly, constrained Pareto efficiency is also studied, e.g., in the context of allocation of resources, stable (or fair) matchings that are not Pareto dominated by other stable (or fair) matchings.

[4] Resoluteness has been a standard property in social choice since its conception and its failure is at the core of the Condorcet paradox, see e.g. Black (1948) and Campbell and Kelly (2003). See Austen-Smith and Banks (1999) for the role of resoluteness in political science, and Zwicker (2016) for a recent survey of canonical social choice results such as Gibbard (1973)-Satterthwaite (1975) Theorem that implicitly or explicitly involve resoluteness.

ences, matching, and allocation of discrete resources without compensating transfers; for earlier uses of the richness assumption we study see Pycia and Troyan (2019).

We analyze welfare criteria imposed on social choice functions and social welfare functions. For every profile of individual preference rankings, a social choice function (SCF) determines what unique alternative should be implemented, while social welfare function (SWF) determines a societal ranking of alternatives. Allowing for partial societal rankings, we can treat an SCF as an SWF in which the outcome of SCF is ranked above all other alternatives.[5] Following Arrow (1963), we say that an SWF is Arrovian if, and only if, it satisfies the standard resoluteness, (strong) Pareto, and independence-of-irrelevant-alternatives postulates. An SWF is resolute if it has a unique social maximum for every profile of preferences; in particular, every SCF is resolute. An SWF satisfies the (strong) Pareto postulate if two socially and Pareto-comparable matchings are ranked so that the Pareto-dominant matching is ranked above the Pareto-dominated one. An SWF satisfies the independence of irrelevant alternatives if, given any two profiles of preferences and any two alternatives that are socially comparable under both profiles, if all individuals rank the two alternatives in the same way under both profiles, then the social ranking of the two alternatives is the same under both profiles. When we want to highlight the positive rather than normative aspects of an SCF we refer to it as a mechanism; we allow here both Arrovian and not Arrovian SCFs. We call a mechanism efficient with respect to an SWF if, for every preference profiles, the resulting outcome is a maximum of the SWF.[6] We say that a mechanism is Arrovian efficient if it is efficient with respect to some Arrovian SWF. Finally, we say that a mechanism is strategy-proof if, for any reports by other individuals, reporting her true ranking leads to the mechanism outcome being weakly better for an individual than any other report.

We introduce a mild auditability requirement that says that, in order to falsify a proposed mechanism outcome, it is sufficient to verify pairwise comparison of individuals' preferences of the outcome with only one challenging alternative (the challenger). This auditability property is attractive as it allows to falsify the mechanism outcome with a limited amount of information and thus largely preserves the privacy of participants' private information.[7]

In Theorem 1, we show that Arrovian efficiency is equivalent to Pareto efficiency and auditability. In Theorem 2 we show that auditability implies non-bossiness of Satterthwaite and Sonnenschein (1981) and in general the reverse implication fails via an example. We prove that the conjunction of individual strategy-proofness and non-bossiness is equivalent to group strategy-proofness, which is in turn equivalent to monotonicity (Maskin, 1999) (Theorem 3).[8] We also

[5] For analysis of welfare with partial orderings, see e.g. see Sen (1970, 1999), Weymark (1984), and Curello and Sinander (2020).

[6] There is a rich social choice literature on the correspondence between choice and the maximum of the SWF ranking in the context of social choice (see below). This literature is interested in rationalizing social choice rather than the efficiency of mechanisms, and hence it talks about mechanisms "rationalized by an SWF" rather than "efficient with respect to an SWF."

[7] For the literature on privacy in mechanism design see the recent survey Pai and Roth (2018).

[8] Analogous two equivalences were established earlier for object allocation, see Pápai (2000) and Takamiya (2001); our proof approach is different and simpler.

show that for Pareto efficient mechanisms, either of these equivalent conditions implies Arrovian efficiency.

We illustrate these results by applying them to characterizations in two canonical economic domains. In voting with the universal strict preference domain, our results immediately imply that Arrovian efficiency and Pareto efficiency are equivalent conditions for an individually strategy-proof mechanism as all mechanisms in the universal domain are non-bossy. In allocation of objects for individuals with unit demand who have strict preferences over the objects—often referred to as house allocation problems—our insights allow us to leverage the results of Pycia and Ünver (2017) to fully characterize the class of auditable and efficient mechanisms as the class of *trading cycles mechanisms*. This characterization provides a no-transfer counterpart of Akbarpour and Li (2020) insight that classical auctions are the "credible" mechanisms in their sense.[9]

We further use this last characterization to show that *almost sequential dictatorships* are the only mechanisms that are individually strategy-proof and Arrovian efficient with respect to a complete SWF, i.e., one that ranks all alternatives, which are matchings in the discrete allocation domain (Theorems 4 and 5). An almost sequential dictatorship combines the ideas of *sequential dictatorship* and *majority voting* between only two possible outcomes. Dictatorships are the benchmark strategy-proof and efficient mechanisms in many areas of economics. When there are three or more alternatives, Gibbard (1973) and Satterthwaite (1975) have shown that all strategy-proof and unanimous voting mechanisms are dictatorial.[10] With two alternatives there are other mechanisms that are strategy-proof and unanimous; majority voting being the primary example. Our class of almost sequential dictatorships combines both of these special mechanisms. Despite these parallels, we find it surprising that the almost sequential dictatorship theorem is true in our environment because—in stark contrast to the environments where this question was previously studied—ours allows many individually strategy-proof (and even group strategy-proof) and Pareto-efficient mechanisms that are not dictatorial.

The present paper is the first to connect the literature on allocation and exchange of discrete resources and the literature on Arrovian preference aggregation. In particular, we seem to be the first to recognize the equivalence of Theorem 2. At the same time, there is a rich literature extending Arrow's program to economic domains, which focuses on determining the class of preference domains in which Arrow's result holds, i.e., economic domains in which all complete Arrovian SWFs are dictatorial; see e.g. Kalai et al. (1979) and Le Breton and Weymark (2011).[11] In addition to us going beyond the dictatorship question, another important difference between this literature and our work is that the earlier literature relies on the weak Pareto postulate as its efficiency concept, that is they say that an alternative is Pareto dominated only if all agents strictly prefer

[9]See also Woodward (2020) for an analysis of a more general concept of auditability in multi-unit auctions.

[10]Dasgupta, Hammond and Maskin (1979) extended this result to more general social choice models. Satterthwaite and Sonnenschein (1981) extended it to public goods economies with production. Zhou (1991) extended it to pure public goods economies. In exchange economies, Barberà and Jackson (1995) showed that strategy-proof mechanisms are Pareto inefficient.

[11]See also Bordes et al. (1995); Bordes and Le Breton (1989, 1990b,a).

another alternative. In contrast, we rely on the more commonly used strong Pareto postulate in economics, in which an alternative is Pareto dominated as soon as all agents weakly prefer another alternative and at least one agent's preference ranking is strict.

Our paper also contributes to the literature on characterizations of dominant strategy mechanisms for house allocation. Ehlers (2002) characterizes group-strategy-proof and Pareto-efficient mechanisms in a maximal domain of weak preferences for which such mechanisms exist and proves a general impossibility result for the domain of all weak preferences.[12] Note that our concept of partial social ranking is different from Ehlers' allowing only certain weak preferences over assigned houses; Ehlers' work is not concerned with social rankings of outcomes and we have equivalence classes for indifferences. Pycia and Ünver (2017) characterizes group-strategy-proof and Pareto-efficient mechanisms in the standard domain of strict preferences and Root and Ahn (2020) characterize properties of these mechanisms allowing for constraints and providing a synthetic treatment of many social choice domains; see also Barberà (1983) and Pápai (2000) who laid the foundations for this line of research. Ma (1994) characterized the class of strategy-proof, individually rational, and Pareto-efficient mechanisms, and his characterization has been extended by Pycia and Ünver (2017) and Tang and Zhang (2015) to richer single-unit demand, by Pápai (2007) to multi-unit demand models, and by Pycia (2016) to settings with network constraints.

Sequential dictatorships have not been studied extensively with unit demand for goods, although their special cases have been. In a *serial dictatorship* (also known as a *priority mechanism*), the same individual chooses next regardless of which house the current individual picks. Svensson (1994) formally introduced and studied serial dictatorships first; Abdulkadiroğlu and Sönmez (1998) studied a probabilistic version of them where the order of individuals is determined uniformly randomly; Svensson (1999) and Ergin (2000) characterized them using plausible axioms. Allowing for outside options, Pycia and Ünver (2007) characterized a subclass of sequential dictatorships different from serial dictatorships. With multiple-house demand under responsive preferences, Hatfield (2009) showed that sequential dictatorships are the only strategy-proof, non-bossy, and Pareto-efficient mechanisms, and Pápai (2001) characterized the sequential dictatorships through the properties of strategy-proofness, non-bossiness, and citizen sovereignty (see also Klaus and Miyagawa, 2002). In a general model allowing both the cases with and without transfers, Pycia and Troyan (2019) showed that a broad class closely resembling sequential dictatorships are precisely the mechanisms that are strongly obviously strategy-proof in their sense; see also Li (2015) and Pycia (2019). For characterizations of random serial dictatorships in terms of incentives, efficiency, and fairness see Liu and Pycia (2011) and Pycia and Troyan (2019). Root and Ahn (2020) characterize the constrained social choice domains in which generalized sequential dictatorships are the only group strategy-proof and Pareto-efficient mechanisms. As an application of their general theorem, they characterize sequential dictatorships as the only mechanisms which are group strategy-proof and Pareto efficient in the roommates problem.

[12] Most of the literature on house allocation—including our paper—is not affected by Ehlers' impossibility result because it analyzes environments in which individuals' preferences are strict.

2 Model

2.1 Environments

Let I be a set of **individuals** and $\mathcal{A}$ be a set of social **alternatives**. Each individual i has a **preference relation** over $\mathcal{A}$ (i.e., a complete, reflexive, and transitive binary relation) denoted by $\succcurlyeq_i$. We denote its strict (i.e., anti-symmetric) part by $\succ_i$ and indifference (i.e., symmetric) part by $\sim_i$. Let $\mathbf{P}_i$ be the domain of preference relations for individual i, and let $\mathbf{P}_J$ denote the Cartesian product $\times_{i \in J} \mathbf{P}_i$ for any $J \subseteq I$. Any profile $\succcurlyeq = (\succcurlyeq_i)_{i \in I}$ from $\mathbf{P} = \mathbf{P}_I$ is called a **preference profile**. For every $\succcurlyeq \in \mathbf{P}$ and $J \subseteq I$, let $\succcurlyeq_J = (\succcurlyeq_i)_{i \in J} \in P_J$ be the restriction of $\succcurlyeq$ to J. Suppose that for every individual there is an exogenous **equivalence relation** $\equiv_i$ on alternative set $\mathcal{A}$. We say that the domain $\mathbf{P}_i$ is **rich** if the following two conditions are satisfied:

1. If for any two alternatives a and b we have $a \equiv_i b$, then for every $\succcurlyeq_i \in \mathbf{P}_i$ we have $a \sim_i b$.
2. If no alternatives in $\mathcal{A}' \subseteq \mathcal{A}$ are $\equiv_i$-equivalent, then all strict preferences on $\mathcal{A}'$ belong to $\mathbf{P}_i$.

Thus, effectively, $\mathbf{P}_i$ is the universal strict preference domain respecting $\equiv_i$-equivalence classes.[13] We say that the preference profile domain $\mathbf{P}$ is **rich** if $\mathbf{P}_i$ is a rich preference domain for every $i \in I$ and for any two alternatives a and b such that $a \equiv_i b$ for every $i \in I$, $a = b$. The last condition eliminates redundancies in our description of the preferences over alternatives. For instance, in house allocation, each social alternative a is a matching between individuals and objects from some set and $a \equiv_i b$ if, and only if, the object matched to i is the same under a and b. In the rest of the paper, we assume that $\mathbf{P}$ is a rich preference profile domain for a fixed equivalence relation profile $(\equiv_i)_{i \in I}$.

Throughout the paper, we fix I and $\mathcal{A}$, and thus, a problem is identified with its preference profile.

A (**direct**) **mechanism** or a **social choice function** (**SCF**) is a mapping $\varphi : \mathbf{P} \to \mathcal{A}$ that assigns an alternative for every preference profile (or, equivalently, for every problem). We denote the outcome of mechanism φ for a preference profile $\succcurlyeq$ as $\varphi[\succcurlyeq]$.

We denote by P^S the set of strict partial orderings over alternatives, where a strict partial ordering is a binary relation that is anti-symmetric and transitive, but not necessarily complete. We refer to elements of P^S as **social rankings**. A **social welfare function** (**SWF**) $\Phi : \mathcal{A} \to P^S$ maps individuals' preference profiles to social rankings. If an alternative a is ranked higher than some other alternative b under $\Phi(\succcurlyeq)$, we denote this as $a\ \Phi(\succcurlyeq)\ b$. An SWF Φ is **resolute** if, for every preference profile $\succcurlyeq$ there exists an alternative a such that $a\ \Phi(\succcurlyeq)\ b$ for every $b \in \mathcal{A} - \{a\}$. We assume SWFs we consider are resolute. A mechanism can be identified with a special instance of

[13] This richness concept was introduced by Pycia and Troyan (2019) who studied it for exogenous structural preference relations (trumping) that can be but not necessarily are equivalence relations. In their terminology, our setting corresponds to no-transfer environments.

a resolute SWF in which the mechanism outcome is the unique maximal alternative of the SWF and no comparisons between non-maximal alternatives are made.

2.2 Efficiency, Auditability, Strategy-Proofness, and Other Properties

An alternative is Pareto efficient for a preference profile $\succcurlyeq$ if no other alternative would make everybody weakly better off and at least one individual better off; that is, an alternative a is **Pareto efficient** if there exists no alternative b such that for every $i \in I$, $b \succcurlyeq_i i$, and for some $i \in I$, $b \succ_i i$. In particular, a mechanism is **Pareto efficient** if it finds a Pareto-efficient alternative for every problem. Pareto efficiency is a weak efficiency requirement and, as discussed in the Introduction, Arrow criticized it for its failure to uniquely determine the best outcome; that is, for not being resolute.[14]

An SWF Φ satisfies the **Pareto** postulate (or is **unanimous**) if: for every preference profile $\succcurlyeq$ and any two alternatives a and b that are comparable by $\Phi(\succcurlyeq)$, if $a \succcurlyeq_i b$ for every $i \in I$, with at least one strict preference, then $a\ \Phi(\succcurlyeq)\ b$.

An SWF Φ satisfies the **independence of irrelevant alternatives (IIA)** if: for every $\succcurlyeq, \succcurlyeq' \in \mathbf{P}$ and $a, b \in \mathcal{A}$,

1. if all individuals rank a and b in the same way, i.e., for every $i \in I$, $a \succcurlyeq_i b \iff a \succcurlyeq'_i b$, and
2. both $\Phi(\succcurlyeq)$ and $\Phi(\succcurlyeq')$ compare a and b, i.e., (i) $a\ \Phi(\succcurlyeq)\ b$ or $b\ \Phi(\succcurlyeq)\ a$ and (ii) $a\ \Phi(\succcurlyeq')\ b$ or $b\ \Phi(\succcurlyeq')\ a$,

then $a\ \Phi(\succcurlyeq')\ b \iff a\ \Phi(\succcurlyeq)\ b$.

We say that an alternative a is **efficient with respect to social ranking** $\succ^S \in P^S$ if it maximizes the social welfare, that is $a \succ^S b$ for every $b \in \mathcal{A} - \{a\}$. A mechanism φ is **efficient with respect to an SWF** Φ if for any profile of individuals' preferences $\succcurlyeq$, the alternative $\varphi[\succcurlyeq]$ is efficient with respect to $\Phi(\succcurlyeq)$. If φ is efficient with respect to some SWF that satisfies the Arrovian postulates of resoluteness, Pareto, and IIA, then we say that φ is **Arrovian efficient**. The next section offers two examples illustrating the concept of Arrovian efficiency.

A mechanism φ is (**one-comparison**) **auditable** if, (i) for any preference profile $\succcurlyeq$, (ii) for any alternative a such that $\varphi[\succcurlyeq] \not\succcurlyeq_j a$ for some individual j, and (iii) for any other preference profile $\succcurlyeq'$ such that the comparisons of alternatives a and $\varphi[\succcurlyeq]$ are the same under $\succcurlyeq_i$ and $\succcurlyeq'_i$ for every $i \in I$, i.e., $\varphi[\succcurlyeq] \succcurlyeq_i a \iff \varphi[\succcurlyeq] \succcurlyeq'_i a$, there exists some individual j' such that $\varphi[\succcurlyeq'] \not\succcurlyeq_{j'} a$. This concept captures the idea that, in order to falsify a proposed alternative as being the outcome of the mechanism, it is sufficient to find one challenger alternative and to verify the pairwise comparisons of the proposed outcome with the challenger. We can thus falsify an outcome with

[14] Notice also that Pareto dominance is a non-resolute SWF.

a limited amount of information; one of the reasons this is an attractive feature of a mechanism is that it allows challenges that rely on relatively little information and largely preserve individuals' privacy.

A mechanism is individually strategy-proof if for every individual, she weakly prefers the outcome when she is truthful to the outcome under any untruthful revelation of her preferences. Formally, a mechanism φ is **individually strategy-proof** if for every $\succcurlyeq \in \mathbf{P}$, there exists no $i \in I$ and $\succcurlyeq_i' \in \mathbf{P}_i$ such that

$$\varphi[\succcurlyeq_i', \succcurlyeq_{-i}] \ \succ_i \ \varphi[\succcurlyeq].$$

A mechanism is non-bossy (Satterthwaite and Sonnenschein, 1981) if when the mechanism chooses two alternatives that are in the same equivalence class of an individual in any two problems that only differ by this individual's preferences, these two alternatives should also be in the same equivalence class of all individuals. Formally, a mechanism is **non-bossy** if for any individual i and for every $\succcurlyeq_i, \succcurlyeq_i' \in \mathbf{P}_i$ and $\succcurlyeq_{-i} \in \mathbf{P}_{-i}$,

$$\varphi[\succcurlyeq_i, \succcurlyeq_{-i}] \equiv_i \varphi[\succcurlyeq_i', \succcurlyeq_{-i}] \implies \varphi[\succcurlyeq_i, \succcurlyeq_{-i}] = \varphi[\succcurlyeq_i', \succcurlyeq_{-i}].$$

A mechanism is group strategy-proof if there is no group of individuals that can misstate their preferences in a way such that each one in the group is weakly better off and at least one individual in the group is strictly better off, irrespective of the preference ranking of the individuals not in the group. Formally, a mechanism φ is **group strategy-proof** if for every $\succcurlyeq \in \mathbf{P}$, there exists no $J \subseteq I$ and $\succ_J' \in \mathbf{P}_J$ such that

$$\varphi[\succcurlyeq_J', \succcurlyeq_{-J}] \ \succcurlyeq_j \ \varphi[\succcurlyeq] \text{ for every } j \in J,$$

and

$$\varphi[\succcurlyeq_J', \succcurlyeq_{-J}] \succ_i \varphi[\succcurlyeq] \text{ for some } i \in J.$$

Given a mechanism φ, a preference profile $\succcurlyeq'$ is a **φ-monotonic transformation** of another preference profile $\succcurlyeq$ if

$$\{a \in \mathcal{A} : a \succcurlyeq_i \varphi[\succcurlyeq]\} \supseteq \{a \in \mathcal{A} : a \succcurlyeq_i' \varphi[\succcurlyeq]\} \text{ for every } i \in I.$$

Thus, for every individual, the set of alternatives weakly better than the mechanism's outcome under the base profile weakly shrinks when we go from the base profile to its monotonic transformation. A mechanism φ is (**Maskin**) **monotonic** (Maskin, 1999) if, for every $\succcurlyeq \in \mathbf{P}$, $\varphi[\succcurlyeq'] = \varphi[\succcurlyeq]$ for every $\succcurlyeq' \in \mathbf{P}$ that is a φ-monotonic transformation of $\succcurlyeq$.

3 Equivalences

In this section, we study individually strategy-proof and Arrovian efficient mechanisms and establish for them equivalence results involving Pareto efficiency, auditability, group strategy-proofness and more technical properties of non-bossiness and monotonicity.

First, we characterize Arrovian efficiency with the help of auditability.[15]

Theorem 1. *A mechanism is Arrovian efficient if, and only if, it is Pareto efficient and auditable.*

Second, auditability is a strictly stronger condition than non-bossiness, even for a Pareto efficient mechanism.

Theorem 2. *Any auditable mechanism is non-bossy. The converse does not hold – even for Pareto-efficient mechanisms.*

Third, the conjunction of the two non-cooperative properties: individual strategy-proofness and non-bossiness is equivalent to either group strategy-proofness or monotonicity. [16]

Theorem 3. *The following three conditions are equivalent for a mechanism:*

1. *group strategy-proofness,*
2. *the conjunction of individually strategy-proofness and non-bossiness,*
3. *monotonicity.*

This result generalizes similar results due to Pápai (2000) and Takamiya (2001) for house allocation environments to our more general setting. Its proof is relegated to the appendix.

To illustrate the results and our concepts, let us look at the *house allocation setting* with three individuals 1, 2, and 3, three houses A, B, and C, and no outside options. Given an alternative, which is a matching of houses to individuals, a let $a(i)$ refer to the house assigned to an individual i under a. Individuals' preferences are denoted by strict preferences over houses, by slight abuse of notation, instead of alternatives and their equivalence relations are over matchings that match them with the same house.[17] In the Appendix, we give an example of a more elaborate incomplete Arrovian SWF, here let us consider two examples of mechanisms illustrating the conditions we study.

[15] In fact, our proof shows something more: for the mechanisms we study, auditability (or non-bossiness) together with Pareto efficiency is also equivalent to Arrovian efficiency with respect to an SWF in which if alternative a Pareto dominates alternative a' then these two alternatives are comparable.

[16] Both of these properties are non-cooperative in the sense that they relate mechanism's outcomes under two scenarios when a single individual makes unilateral preference-revelation deviations.

[17] We formally define this setting in the next section.

Example 1: With three individuals $1, 2, 3$ and three houses A, B, C (thus, with 6 alternatives) the serial dictatorship φ in which individual 1 chooses first the house she would like to receive and individual 2 chooses second is well-known to be individually strategy-proof, non-bossy, and Pareto efficient, as well as group strategy-proof and monotonic.

It is straightforward to see that this serial dictatorship is Arrovian efficient with respect to the following SWF: a is ranked above b if and only if (a) 1 prefers a to b, or (b) 1 is indifferent and 2 prefers a to b.

As φ treats all objects in a symmetric (neutral) way, to establish the serial dictatorship's auditability, it is sufficient to look at a preference profile $\succcurlyeq$ such that $\varphi[\succcurlyeq] = \{(1, A), (2, B), (3, C)\}$, a different alternative b and any preference profile $\succcurlyeq'$ such that $\succcurlyeq'_i$ keeps the same ranking as $\succcurlyeq_i$ between $\varphi[\succcurlyeq]$ and b for every individual i and to show that $\varphi[\succcurlyeq'] \neq b$. To verify this inequality consider two cases:

- $A \neq b(1)$. Then $A \succ_1 b(1)$ because 1 being the first dictator chose her top choice under $\succcurlyeq_i$. Hence, $A \succ'_1 b(1)$. 1 is not choosing $b\,(1)$ when having preference ranking $\succ'_1$ and thus $\varphi[\succcurlyeq'] \neq b$.
- $A = b(1)$. Then either :
  - ⋆ $B \neq b(2)$. Then, $B \succ_2 b(2)$ by an argument similar to the previous case. If $\varphi[\succcurlyeq'](1) = B$ then $\varphi[\succcurlyeq'] \neq b$, and the auditability inequality obtains. If $\varphi[\succcurlyeq'](1) \neq B$ then either $\varphi[\succcurlyeq'](1) \neq A = b(1)$ and the auditability inequality obtains, or $\varphi[\succcurlyeq'](1) = A = b(1)$ and hence B is available when 2's assignment is determined, and thus, $\varphi[\succcurlyeq'] \succcurlyeq'_2 B \succ'_2 b(2)$, and hence, $\varphi[\succcurlyeq'] \neq b$ and the auditability inequality obtains.
  - ⋆ $B = b(2)$. Then $C = b(3)$ contrary to $b \neq \varphi[\succcurlyeq]$.

■

Example 2: We now modify the serial dictatorship of the previous example and consider mechanisms ψ in which 1 chooses first; then 2 chooses second if 1 prefers B over C, else 3 chooses second. This mechanism is an example of a ranking-dependent sequential dictatorship, and is also individually strategy-proof and Pareto efficient. However, mechanism ψ is neither Arrovian efficient nor non-bossy nor auditable. To see the latter three points, let us look at the following two preference profiles, which differ only in how individual 1 ranks objects:

$$\succcurlyeq = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & A & A \\ B & B & B \\ C & C & C \end{array} \quad \succcurlyeq' = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & A & A \\ C & B & B \\ B & C & C \end{array},$$

and notice that

$$\begin{aligned}\psi[\succcurlyeq] &= \{(1,A),(2,B),(3,C)\}, \\ \psi[\succcurlyeq'] &= \{(1,A),(2,C),(3,B)\}.\end{aligned}$$

Mechanism ψ does not satisfy non-bossiness because from $\succcurlyeq$ to $\succcurlyeq'$ only 1's preference changes and her assignment does not change, and yet other individuals' assignments change (leading to different equivalence classes of alternatives for either individual 2 and 3).

Mechanism ψ does not satisfy Arrovian efficiency. Indeed, by way of contradiction assume that ψ is Arrovian efficient with respect to some Arrovian SWF Ψ. Then $\Psi(\succcurlyeq)$ ranks alternative $\psi[\succcurlyeq]$ above $\psi[\succcurlyeq']$, and $\Psi(\succcurlyeq')$ ranks $\psi[\succcurlyeq']$ above $\psi[\succcurlyeq]$. But, this violates IIA, a contradiction that shows that ψ is not Arrovian efficient.

Mechanism ψ does not satisfy auditability as we can contest the alternative $\psi[\succcurlyeq]$ with alternative $b = \psi[\succcurlyeq']$.

Mechanism ψ does not satisfy group strategy-proofness because the group $\{1,3\}$ can beneficially manipulate by reporting $\succcurlyeq'_{\{1,3\}}$ instead of $\succcurlyeq_{\{1,3\}}$ (noticing $\succcurlyeq_2 = \succcurlyeq'_2$), making individual 3 strictly better off while leaving individual 1 indifferent.

Finally, mechanism ψ does not satisfy monotonicity as $\succcurlyeq'$ is a ψ-monotonic transformation of $\succcurlyeq$ and yet the mechanism's respective outcomes are in different equivalence classes for individuals 2 and 3. ■

We are ready to prove Theorems 1 and 2.

Proof of Theorem 1.

(Arrovian efficiency $\Longrightarrow$ Pareto efficiency) Consider an Arrovian efficient mechanism φ with respect to some SWF Φ. Suppose that for some $\succcurlyeq \in \mathbf{P}$, $\varphi[\succcurlyeq]$ is not Pareto efficient. Then there exists some $a \in \mathcal{A} - \{\varphi[\succcurlyeq]\}$ such that $a \succcurlyeq_i \varphi[\succcurlyeq]$ for every i, with a strict preference for at least one individual. Because Φ satisfies the Pareto postulate, we have $a\, \Phi(\succcurlyeq)\, \varphi[\succcurlyeq]$, which contradicts the assumption that φ is Arrovian efficient with respect to Φ.

(Arrovian efficiency $\Longrightarrow$ auditability). An inspection of the definitions shows that Arrovian efficiency directly implies auditability; indeed, auditability is effectively IIA restricted to comparisons involving the top equivalence class.

(Pareto efficiency and auditability $\Longrightarrow$ Arrovian efficiency). Consider a Pareto-efficient and auditable mechanism φ. We define an SWF Φ as follows: for any profile of preferences $\succcurlyeq$ and any two alternatives a and $a' \neq a$, alternative a is ranked by $\Phi(\succcurlyeq)$ above a' if, and only if, either (i) we have $a = \varphi[\succcurlyeq]$ or (ii) for every individual i, we have $a \succcurlyeq_i a'$ and at least for one individual i the preference is strict (which we refer to, by sight abuse of terminology, as "individuals unanimously rank a over a'" throughout the proof). Note that Pareto efficiency of φ implies that conditions (i) and (ii) are consistent with each other, and hence, that the SWF Φ is well defined.

By definition, Φ satisfies the Pareto postulate. Furthermore, Φ is transitive: if $\Phi(\succcurlyeq)$ ranks a^1 above a^2 and it ranks a^2 above a^3 , then it ranks a^1 above a^3. To see this: if one of these a^ℓ (for $\ell = 1, 2, 3$) equals $\varphi[\succcurlyeq]$, then it must be that $a^1 = \varphi[\succcurlyeq]$, and the claim is proven. If none of the a^ℓ equals $\varphi[\succcurlyeq]$, then individuals unanimously rank a^1 above a^2 and unanimously rank a^2 above a^3; we conclude that individuals unanimously rank a^1 above a^3, and thus, $\Phi(\succcurlyeq)$ ranks a^1 above a^3 by construction.

It remains to check that Φ satisfies IIA. Take two preference profiles $\succcurlyeq^1$ and $\succcurlyeq^2$ such that each individual ranks two alternatives, say a and a', in the same way under the two preference profiles. If the two alternatives are comparable under both $\Phi\left(\succcurlyeq^1\right)$ and $\Phi\left(\succcurlyeq^2\right)$, then one of the following cases obtains:

Case 1: One of the alternatives is unanimously preferred to the other under $\succcurlyeq^1$; then the same unanimous preference obtains under $\succcurlyeq^2$ and the claim is true.

Case 2: There is no unanimous preference of the two alternatives under $\succcurlyeq^1$; then unanimity cannot obtain under $\succcurlyeq^2$ either. As the alternatives are ranked, it must be that $\varphi[\succcurlyeq^1], \varphi[\succcurlyeq^2] \in \{a, a'\}$ by construction of Φ. Suppose, without loss of generality $\varphi[\succcurlyeq^1] = a$. If $a \equiv_i a'$ for all $i \in I$ then $a = a'$ by richness assumption. So suppose for some individual i, $a' \not\equiv_i a = \varphi[\succcurlyeq^1]$. As all individuals rank $\varphi[\succcurlyeq^1] = a$ and a' the same way under $\succcurlyeq^1$ and $\succcurlyeq^2$, and φ is auditable then $\varphi[\succcurlyeq^2] \neq a'$. This implies $\varphi[\succcurlyeq^2] = a$, as well. Since φ always picks the unique top alternative of the SWF Φ, then $a \ \Phi(\succcurlyeq^1) \ a'$ and $a \ \Phi(\succcurlyeq^2) \ a'$. Thus, Φ satisfies IIA. QED

Proof of Theorem 2. To show that auditable φ is non-bossy, let $\succcurlyeq \in \mathbf{P}$ and, for an individual i, $\succcurlyeq'_i \in \mathbf{P}_i$ be such that

$$\varphi[\succcurlyeq] \equiv_i \varphi[\succcurlyeq'_i, \succcurlyeq_{-i}].$$

Suppose, by way of contradiction, that there is an individual j for whom $\varphi[\succcurlyeq'_i, \succcurlyeq_{-i}] \not\equiv_j \varphi[\succcurlyeq]$. This contradicts auditability between alternative $a = \varphi[\succcurlyeq'_i, \succcurlyeq_{-i}]$ and $\varphi[\succcurlyeq]$ because all individuals rank alternatives a and $\varphi[\succcurlyeq]$ in the same way under $\succcurlyeq$ and $(\succcurlyeq'_i, \succcurlyeq_{-i})$; yet, $\varphi[\succcurlyeq'_i, \succcurlyeq_{-i}] = a$. Thus, for every $j \in I$, $\varphi[\succcurlyeq'_i, \succcurlyeq_{-i}] \equiv_j \varphi[\succcurlyeq]$, which in turn implies by richness assumption $\varphi[\succcurlyeq'_i, \succcurlyeq_{-i}] = \varphi[\succcurlyeq]$.

To show that non-bossiness does not in general imply auditability even for Pareto-efficient mechanisms, consider the voting environments with the universal strict preference domain over alternatives. In this domain, every mechanism is non-bossy by definition. Suppose $\mathcal{A} = \{a_1, a_2, \ldots, a_k\}$ for some $k \geq 3$. Consider the *plurality voting mechanism* φ which selects as the outcome the alternative that is ranked as the top choice by most individuals (and if there are multiple such alternatives then chooses the one with the smallest index[18]). This mechanism is clearly Pareto efficient but not

[18]Formally, let $v_a(\succcurlyeq) = |\{i \in I : a \in \max_{\succcurlyeq_i} \mathcal{A}\}|$ be the number of *votes* an alternative a gets under a preference profile $\succcurlyeq$; then $\varphi[\succcurlyeq] \in \arg\max_{a \in \mathcal{A}} v_a(\succcurlyeq)$ with the property that $\varphi[\succcurlyeq]$ has the smallest index among all alternatives in the set $\arg\max_{a \in \mathcal{A}} v_a(\succcurlyeq)$.

auditable. To see the last point consider two preference profiles $\succcurlyeq$ and $\succcurlyeq'$ with three individuals $I = \{1, 2, 3\}$ and three alternatives $\mathcal{A} = \{a_1, a_2, a_3\}$:

$$\succcurlyeq = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline a_1 & a_2 & a_3 \\ \vdots & a_3 & \dots \\ & a_1 & \end{array} \qquad \succcurlyeq' = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline a_1 & a_3 & a_3 \\ \vdots & \vdots & \vdots \\ & & \end{array},$$

We have $\varphi[\succcurlyeq] = a_1$ as all alternatives receive one vote and a_1 has the lowest index, while $\varphi[\succcurlyeq'] = a_3$ with the highest votes. On the other hand, $\succcurlyeq$ and $\succcurlyeq'$ rank relatively a_1 and a_3 the same (individual 1 prefers a_1 while 2 and 3 prefer a_3) yet $\varphi[\succcurlyeq'] = a_3 \not\equiv_i \varphi[\succcurlyeq]$ for any $i \in I$. Thus, φ is not auditable. QED

The results of the current section have the following immediate corollaries:[19]

Corollary 1. *A mechanism is Arrovian efficient and individually strategy-proof if, and only if, it is Pareto efficient and group strategy-proof if, and only if, it is Pareto efficient and monotonic.*

Corollary 2. *Suppose that a mechanism is individually strategy-proof and Pareto efficient. Then, the following five conditions are equivalent for this mechanism:*

- *Arrovian efficiency,*
- *auditability,*
- *group strategy-proofness,*
- *non-bossiness,*
- *monotonicity.*

4 Applications

4.1 Auditability, Pareto Efficiency, and Arrovian Efficiency in Voting

The most straightforward application of result is in the universal strict preference domain, which is also often called the universal *voting environment*. This environment consists of all strict preference relations over alternatives, each of which can be interpreted as a candidate in an election.

[19]It is worthy to note that Arrovian efficiency does not in general imply individual strategy-proofness: Consider an environment with at least 3 individuals $I \supseteq \{1, 2, 3\}$ and 2 alternatives $\mathcal{A} = \{a, b\}$. Suppose individuals' preferences consist of strict preferences over these two alternatives. Consider the following mechanism: If individual 3 ranks a higher than b then individual 1's top choice is implemented, and otherwise individual 2's top choice is implemented. This mechanism is Arrovian efficient while individual 3 can manipulate it.

Corollary 3. *In the universal strict preference domain, for an individually strategy-proof mechanism the following two conditions are equivalent:*

- *Pareto efficiency,*
- *Arrovian efficiency.*

One direction of the corollary follows from Theorem 2 and then Theorem 1 because, in the universal strict preference domain, every mechanism is non-bossy, and the other direction was established in Theorem 1.

4.2 Incomplete and Complete SWFs in House Allocation

We now apply our results to house allocation problems. Formally, a **house allocation** environment consists of the set of individuals I and a set of **houses** $\mathcal{H}$. A social alternative for this problem is a matching. To simplify the definition of a matching, we focus on environments in which $|\mathcal{H}| \geq |I|$. To define a matching, let us start with a more general concept that we use frequently below. A submatching is an allocation of a subset of houses to a subset of individuals, such that no two different individuals get the same house. Formally, a **submatching** is a one-to-one function $s : J \to \mathcal{H}$; where for $J \subseteq I$, using the standard function notation, we denote by $s(i)$ the assignment of individual $i \in J$ under s, and by $s^{-1}(H)$ the individual that got house $H \in s(J)$ under s. Let $\mathcal{S}$ be the set of submatchings. For every $s \in \mathcal{S}$, let I_s denote the set of individuals matched by s and $\mathcal{H}_s \subseteq \mathcal{H}$ denote the set of houses matched by s. For every $H \in \mathcal{H}$, let $\mathcal{S}_{-H} \subset \mathcal{S}$ be the set of submatchings $s \in \mathcal{S}$ such that $H \in \mathcal{H} - \mathcal{H}_s$, i.e., the set of submatchings at which house H is unmatched. By virtue of the set-theoretic interpretation of functions, submatchings are sets of individual-house pairs and are ordered by inclusion. A **matching**, which is the social alternative in this context, is a maximal submatching; that is, $a \in \mathcal{S}$ is a matching if $I_a = I$. As before, let $\mathcal{A} \subset \mathcal{S}$ be the set of matchings. We will write $\overline{I_s}$ for $I - I_s$ and $\overline{\mathcal{H}_s}$ for $\mathcal{H} - \mathcal{H}_s$ for short. We will also write $\overline{\mathcal{A}}$ for $\mathcal{S} - \mathcal{A}$.

For any individual i, her equivalence relation $\equiv_i$ over matchings is defined over the matching in which is assigned the same house: For any $a, b \in \mathcal{A}$, $a \equiv_i b \iff a(i) = b(i)$. Each of her preference relations $\succcurlyeq_i \in \mathbf{P}_i$, which satisfies the richness assumption, has the following property: for any $a, b \in \mathcal{A}$, $a \sim_i b \iff a(i) = b(i)$. With slight abuse of terminology we also use this preference relation to denote her preference relation over houses i.e. for any $H, H' \in \mathcal{H}$, $H \succcurlyeq_i H' \iff a \succcurlyeq_i a'$ for any two matchings a, a' such that $a(i) = H$ and $a'(i) = H'$.

We introduce the full class of individually strategy-proof, non-bossy, and Pareto-efficient mechanisms, as characterized by Pycia and Ünver (2017), which will be used to obtain the main result in this section. This is the class of trading-cycles mechanisms. This mechanism class is defined through an iterative algorithm, which matches some individuals in every round. Depending on who is matched with which house in preceding rounds, the remaining houses are controlled by the remaining individuals in a round of the algorithm. We define a control-rights structure as a function of the submatching that is fixed: A **structure of control rights** is a collection of mappings

$$(\kappa, \beta) = \left\{ (\kappa_s, \beta_s) : \overline{\mathcal{H}_s} \to \overline{I_s} \times \{\text{ownership, brokerage}\} \right\}_{s \in \overline{\mathcal{A}}}.$$

The functions κ_s of the control-rights structure tell us which unmatched individual controls any particular unmatched house at a submatching s, where **at** s is the terminology we use when some individuals and houses are already matched with respect to s. Agent i **controls** house $H \in \overline{\mathcal{H}_s}$ at submatching s when $\kappa_s(H) = i$. The type of control is determined by functions β_s. We say that the individual $\kappa_s(H)$ **owns** H at s if $\beta_s(H) =$ownership, and that the individual $\kappa_s(H)$ **brokers** H at s if $\beta_s(H) =$brokerage. In the former case, we call the individual an **owner** and the controlled house an **owned house**. In the latter case, we use the terms **broker** and **brokered house**. Notice that each controlled (owned or brokered) house is unmatched at s, and any unmatched house is controlled by some uniquely determined unmatched individual. We need to impose certain conditions on the control-rights structures to guarantee that the induced mechanisms are individually strategy-proof, non-bossy, and Pareto efficient.

A structure of control rights (κ, β) is **consistent** if the following within-round and across-round requirements are satisfied for every $s \in \overline{\mathcal{A}}$:

Within-Round Requirements:

(**R1**) There is at most one brokered house at s, or $|\overline{\mathcal{H}_s}| = 3$ and all remaining houses are brokered.

(**R2**) If i is the only unmatched individual at s, then i owns all unmatched houses at s.

(**R3**) If individual i brokers a house at s, then i does not control any other houses at s.

Across-Round Requirements: Consider submatching s' such that $s \subset s' \in \overline{\mathcal{A}}$, and an individual $i \in \overline{I_{s'}}$ that owns a house $H \in \overline{\mathcal{H}_{s'}}$ at s. Then:

(**R4**) Agent i owns H at s'.

(**R5**) If i' brokers house H' at s, and $i' \in \overline{I_{s'}}$, $H' \in \overline{\mathcal{H}_{s'}}$, then either i' brokers H' at s', or i owns H' at s'. (Notice that the latter case can only happen if i is the only individual in $\overline{I_{s'}}$ who owns a house at s.)

(R6) If individual $i' \in \overline{I_{s'}}$ controls $H' \in \overline{\mathcal{H}_{s'}}$ at s, then i' owns H at $s \cup \{(i, H')\}$.

Each consistent control-rights structure (κ, β) induces a **trading-cycles (TC)** mechanism $\psi^{\kappa,\beta}$, and given a problem $\succcurlyeq \in \mathbf{P}$, the outcome matching $\psi^{\kappa,\beta}[\succcurlyeq]$ is found as follows:

The TC algorithm:

The algorithm starts with empty submatching $s^0 = \varnothing$ and in each round $r = 1, 2, ...$ it matches some individuals with houses. By s^{r-1}, we denote the submatching of individuals matched before round r. If $s^{r-1} \in \overline{\mathcal{A}}$, then the algorithm proceeds with the following three steps of round r:

Step 1 *Pointing.* Each house $H \in \overline{\mathcal{H}_{s^{r-1}}}$ points to the individual who controls it at s^{r-1}. Each individual $i \in \overline{I_{s^{r-1}}}$ points to her most preferred outcome in $\overline{\mathcal{H}_{s^{r-1}}}$.

Step 2(a) *Matching Simple Trading Cycles.* A cycle

$$H^1 \to i^1 \to H^2 \to ...H^n \to i^n \to H^1,$$

in which $n \in \{1, 2, ...\}$ and individuals $i^\ell \in \overline{I_{s^{r-1}}}$ point to houses $H^{\ell+1} \in \overline{\mathcal{H}_{s^{r-1}}}$ and houses H^ℓ point to individuals i^ℓ (here $\ell = 1, ..., n$ and superscripts are added modulo n), is simple when at least one individual in the cycle is an owner. Each individual in each simple trading cycle is matched with the house she is pointing to.

Step 2(b) *Forcing Brokers to Downgrade Their Pointing.* If there are no simple trading cycles in the preceding Step 2(a), and only then we proceed as follows (otherwise we proceed to step 3).

- ⋆ If there is a cycle in which a broker i points to a brokered house, and there is another broker or owner that points to this house, then we force broker i to point to her next choice and we return to Step 2(a).[20]
- ⋆ Otherwise, we clear all trading cycles by matching each individual in each cycle with the house she is pointing to.

Step 3 Submatching s^r is defined as the union of s^{r-1} and the set of newly matched individual-house pairs. When all individuals or all houses are matched under s^r, then the algorithm terminates and gives matching s^r as its outcome.

One important feature of the TC mechanisms is that we can, without loss of generality, rule out the existence of brokers at some submatching s if there is a single owner at s. We formalize this property as a remark:

Remark 1. Pycia and Ünver (2017) *For every TC mechanism such that for some s there is only one owner and one broker, there is an equivalent TC mechanism such that at s there are no brokers and the same owner owns all houses.*

[20]Importantly, broker i is unique by R1.

Using Theorem 2 and Pycia and Ünver (2017)'s characterization we obtain the following corollary:

Corollary 4. *A mechanism is individually strategy-proof and Arrovian efficient if, and only if it is a TC mechanism.*

So far, we allowed welfare functions to incompletely rank social outcomes. We now show that a class that we refer to as almost sequential dictatorships is exactly the mechanisms that are strategy-proof and Arrovian efficient with respect to complete SWF, that is SWF that always rank all outcomes.

First we define the following class: a **top-trading-cycles (TTC)** (or **hierarchical exchange**) mechanism is a TC mechanism with a control-rights structure in which no house is ever brokered at any submatching (Pápai, 2000). A TTC mechanism $\psi^{\kappa,\beta}$ will be denoted by dropping β from its notation as ψ^{κ}.

TTC mechanisms form a strict subclass of TC mechanisms. Let us start with a lemma showing that not every TTC is Arrovian efficient with respect to a complete SWF.

Lemma 1. *Suppose that $|\mathcal{H}| \geq |I| = 2$ and a TTC mechanism is Arrovian efficient with respect to a complete SWF. Then in this mechanism no individual can own two houses while a second individual owns a house.*

Proof. Consider allocating three or more houses to two individuals. Let φ be a TTC mechanism in which individual 1 owns house A and individual 2 owns houses B and C. To see that there is no complete SWF such that φ is efficient, consider the preference profile

$$\succcurlyeq = \begin{array}{|c|c|} 1 & 2 \\ \hline B & A \\ A & B \\ C & C \\ \vdots & \vdots \end{array}.$$

and the following four auxiliary preference profiles

$$\succcurlyeq^1 = \begin{array}{|c|c|} 1 & 2 \\ \hline B & C \\ A & \vdots \\ \vdots & \end{array}, \quad \succcurlyeq^2 = \begin{array}{|c|c|} 1 & 2 \\ \hline B & B \\ C & C \\ \vdots & \vdots \end{array}, \quad \succcurlyeq^3 = \begin{array}{|c|c|} 1 & 2 \\ \hline C & A \\ \vdots & B \\ & \vdots \end{array}, \quad \succcurlyeq^4 = \begin{array}{|c|c|} 1 & 2 \\ \hline A & A \\ C & C \\ \vdots & \vdots \end{array}.$$

Denote

$$
\begin{aligned}
a^1 &= \varphi[\succsim^1] = \{(1, B), (2, C)\}, \\
a^2 &= \varphi[\succsim^2] = \{(1, C), (2, B)\}, \\
a^3 &= \varphi[\succsim^3] = \{(1, C), (2, A)\}, \\
a^4 &= \varphi[\succsim^4] = \{(1, A), (2, C)\}.
\end{aligned}
$$

Now, if there is a complete SWF Φ such that φ is Arrovian efficient, then $\Phi\left(\succsim^1\right)$ ranks a^1 above a^4, and by IIA, this implies that $\Phi\left(\succsim\right)$ ranks a^1 above a^4. Similarly, $\Phi\left(\succsim^2\right)$ ranks a^2 above a^1, and by IIA, this implies that $\Phi\left(\succsim\right)$ ranks a^2 above a^1. Further, and again similarly, $\Phi\left(\succsim^3\right)$ ranks a^3 above a^2, and by IIA, this implies that $\Phi\left(\succsim\right)$ ranks a^3 above a^2. Finally, $\Phi\left(\succsim^4\right)$ ranks a^4 above a^3, and by IIA, this implies that $\Phi\left(\succsim\right)$ ranks a^4 above a^3. But then $\Phi\left(\succsim\right)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient. QED

We will use this lemma to characterize individually strategy-proof and Arrovian efficient mechanisms for $|\mathcal{H}| > |I|$; we will characterize this class of mechanisms for $|\mathcal{H}| = |I|$ later. The resulting class consists of sequential dictatorships. Formally, a **sequential dictatorship** is a TTC mechanism ψ^κ such that for every $s \in \overline{\mathcal{A}}$ and $H, H' \in \overline{\mathcal{H}_s}$, $\kappa_H(s) = \kappa_{H'}(s)$, i.e., an unmatched individual owns all unmatched houses at s. For notational convenience, we will represent each $\kappa_H(\cdot)$ as $\kappa(\cdot)$. Sequential dictatorships turn out to be the class of Arrovian-efficient and individually strategy-proof mechanisms for this case:

Theorem 4. *Suppose $|\mathcal{H}| > |I|$. A mechanism is individually strategy-proof and Arrovian efficient with respect to a complete social welfare function if, and only if, it is a sequential dictatorship.*

Proof of Theorem 4. If $|I| = 1$, the theorem is trivially true. Suppose $|I| \geq 2$.

($\Longrightarrow$) Consider a mechanism φ that is individually strategy-proof and efficient with respect to a complete Arrovian welfare function. By Theorem 2 and Corollary 4, φ is a TC mechanism $\psi^{\kappa,\beta}$.

Fix an arbitrary preference profile $\succsim \in \mathbf{P}$. We claim that at any round r of the algorithm $\psi^{\kappa,\beta}$, there is exactly one individual who controls all houses. We prove it in two steps. First, let us show that there cannot be two (or more) individuals who each own a house. By way of contradiction, suppose that some individual 1 controls house A and some other individual 2 controls house B in round r.

Suppose s is the submatching created by the TC algorithm for $\psi^{\kappa,\beta}$ before round r at $\succsim$. Fix house $C \in \{A, B\}$ as an unmatched house at s. Consider four auxiliary preference profiles $\succsim^\ell$ that all share the following properties: (i) each individual matched under s ranks houses under $\succsim^\ell$, $\ell = 1, ..., 4$, in the same way they rank them under $\succsim$, (ii) each individual i unmatched at s and different from individuals 1 and 2 ranks a unique s-unmatched house $H_i \notin \{A, B, C\} \cup \mathcal{H}_s$ as her first choice (such a unique house exists as $|\mathcal{H}| > |I|$), and (iii) individuals 1 and 2 each rank all houses other than A, B, C lower than A, B, C. In particular, the four profiles differ only in how individuals 1 and 2 rank houses A, B, C: the ranking of A, B, C is the same as in the four preference profiles from the proof of Lemma 1 above. Notice that

$$\psi^{\kappa,\beta}[\succsim^{\ell}] = s \cup a^{\ell} \cup \{(i, H_i)\}_{i \in \overline{I_s} - \{1,2\}},$$

where a^{ℓ}s are defined as in the proof of Lemma 1 above. Furthermore, the same argument we used in the proof of the lemma shows that there can be no SWF that ranks all four a^{ℓ}s, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{\kappa,\beta}$ efficient, a contradiction that implies that there cannot be two individuals who own houses in a round of the algorithm.

As $\psi^{\kappa,\beta}$ never allows two owners in a round of the algorithm, by Corollary 4 and Remark 1, without loss of generality we can assume that there are no brokers in any round, either. Hence, in each round of the algorithm there is a single individual who controls (and owns) all houses. That means that $\psi^{\kappa,\beta}$ is a sequential dictatorship.

($\Longleftarrow$) Consider a sequential dictatorship ψ^{κ}. We construct a complete SWF Φ such that ψ^{κ} is efficient with respect to Φ. Under Φ any two matchings are ranked according to the preference relation of the first-round dictator; if she is indifferent , then the matchings are ranked according to the preference relation of the second-round dictator, etc. Formally, for any $\succsim \in \mathbf{P}$ and any two distinct $a, b \in A$, let $a\,\Phi(\succsim)\,\beta$ if and only if there exists $k \in \{1, ..., |I|\}$ such that $a\,(i_1) = b\,(i_1)$, ... and $a\,(i_{k-1}) = b\,(i_{k-1})$, and individual i_k prefers $a\,(i_k)$ over $b\,(i_k)$, where individuals $i_1, ..., i_k$ are defined recursively: $i_1 = \kappa\,(\emptyset)$, and in general $i_{\ell} = \kappa\,(\{(i_1, a\,(i_1))\,, ..., (i_{\ell-1}, a\,(i_{\ell-1}))\})$ for $\ell = 1, ..., k$. It is straightforward to verify that Φ is a complete Arrovian SWF and that ψ^{κ} is efficient with respect to Φ. QED

While the above argument relies on there being more houses than agents, we can modify the argument to characterize the case $|\mathcal{H}| = |I|$. The class of individually strategy-proof and Arrovian efficient mechanisms consists then sequential dictatorships and some additional new mechanisms. An **almost sequential dictatorship** is a TTC mechanism ψ^{κ} such that for every $s \in \overline{A}$ such that $|\overline{\mathcal{H}_s}| \neq 2$ we have $\kappa_H(s) = \kappa_{H'}(s)$ for every $H, H' \in \overline{\mathcal{H}_s}$. Note that the only mechanisms that are not sequential dictatorships in this class are mechanisms that assign to different owners each of the houses when only two houses are left, but otherwise a single individual owns all houses.

Our final result is as follows:

Theorem 5. *A mechanism is individually strategy-proof and Arrovian efficient with respect to a complete SWF if, and only if, it is an almost sequential dictatorship.*

In the proof we use Lemma 1 and two further lemmas showing that three individuals each cannot simultaneously control a house under a TC mechanism that is efficient with respect to a complete SWF.

Lemma 2. *Suppose that $|\mathcal{H}| = |I| \geq 3$ and a TC mechanism is Arrovian efficient with respect to a complete SWF. Then in this mechanism one individual cannot control a house while two others each own a house.*

Proof. Consider a TC mechanism φ in which individual 1 owns house A, individual 2 owns house B, and individual 3 controls house C. We will show that there is no complete SWF such that φ is Arrovian efficient. Consider the preference profile

$$\succcurlyeq = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & C & A \\ C & A & B \\ A & B & C \\ \vdots & \vdots & \vdots \end{array}.$$

and the following three additional preference profiles

$$\succcurlyeq^1 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & C & B \\ C & \vdots & \vdots \\ A & & \\ \vdots & & \end{array}, \quad \succcurlyeq^2 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline C & C & A \\ \vdots & A & \vdots \\ & B & \\ & \vdots & \end{array}, \quad \succcurlyeq^3 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & A & A \\ \vdots & \vdots & B \\ & & C \\ & & \vdots \end{array}.$$

Regardless of whether individual 3 owns or brokers house C, we have

$$\begin{aligned} a^1 &= \varphi[\succcurlyeq^1] = \{(1, A), (2, C), (3, B)\}; \\ a^2 &= \varphi[\succcurlyeq^2] = \{(1, C), (2, B), (3, A)\}; \\ a^3 &= \varphi[\succcurlyeq^3] = \{(1, B), (2, A), (3, C)\}. \end{aligned}$$

If there is a complete SWF Φ such that φ is Arrovian efficient, then $\Phi\left(\succcurlyeq^1\right)$ ranks a^1 above a^3, and by IIA, this implies that $\Phi(\succcurlyeq)$ ranks a^1 above a^3. Similarly, $\Phi\left(\succcurlyeq^2\right)$ ranks a^2 above a^1, and by IIA, this implies that $\Phi(\succcurlyeq)$ ranks a^2 above a^1. Further, and again similarly, $\Phi\left(\succcurlyeq^3\right)$ ranks a^3 above a^2, and by IIA, this implies that $\Phi(\succcurlyeq)$ ranks a^3 above a^2. Then $\Phi(\succcurlyeq)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient. QED

Lemma 3. *Suppose that $|\mathcal{H}| = |I| \geq 3$ and a TC mechanism is Arrovian efficient with respect to a complete SWF. Then, in any round of the TC algorithm, there is at most one broker.*

Proof. By way of contradiction, suppose that in some round of the TC mechanism there are more than one broker and let φ be the continuation TC mechanism from this round onwards. Without loss of generality, in φ individual 1 brokers house A, individual 2 brokers house B, and individual

3 brokers house C. We will show that there is no complete SWF such that φ is Arrovian efficient. Consider the following preference profiles

$$\succcurlyeq = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & B & C \\ A & A & B \\ C & C & A \\ \vdots & \vdots & \vdots \end{array}.$$

and

$$\succcurlyeq^1 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & B & C \\ C & A & B \\ \vdots & \vdots & \vdots \end{array}, \quad \succcurlyeq^2 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & B & C \\ A & C & A \\ \vdots & \vdots & \vdots \end{array}, \quad \succcurlyeq^3 = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline B & A & B \\ C & C & A \\ \vdots & \vdots & \vdots \end{array}.$$

Denote

$$a^1 = \varphi[\succcurlyeq^1] = \{(1, A), (2, B), (3, C)\};$$
$$a^2 = \varphi[\succcurlyeq^2] = \{(1, B), (2, C), (3, A)\};$$
$$a^3 = \varphi[\succcurlyeq^3] = \{(1, C), (2, A), (3, B)\}.$$

If there is a complete SWF Φ such that φ is Arrovian efficient, then $\Phi\left(\succcurlyeq^1\right)$ ranks a^1 above a^3, and by IIA, this implies that $\Phi\left(\succcurlyeq\right)$ ranks a^1 above a^3. Similarly, $\Phi\left(\succcurlyeq^2\right)$ ranks a^2 above a^1, and by IIA, this implies that $\Phi\left(\succcurlyeq\right)$ ranks a^2 above a^1. Further, again similarly, $\Phi\left(\succcurlyeq^3\right)$ ranks a^3 above a^2, and by IIA, this implies that $\Phi\left(\succcurlyeq\right)$ ranks a^3 above a^2. Then $\Phi\left(\succcurlyeq\right)$ fails transitivity, showing that there does not exist a complete SWF with respect to which φ is efficient. QED

Proof of Theorem 5. If $|\mathcal{H}| > |I|$, it follows from Theorem 4 and if $|\mathcal{H}| = |I| = 1$, the theorem is trivially true. Hence, suppose $|\mathcal{H}| = |I| > 1$.

($\Longrightarrow$) Consider a mechanism φ that is individually strategy-proof and efficient with respect to a complete Arrovian welfare function. By Theorem 2 and Corollary 4, φ is a TC mechanism $\psi^{\kappa,\beta}$.

Fix $\succcurlyeq \in \mathbf{P}$. We claim that at any round r of the algorithm for $\psi^{\kappa,\beta}$, there is exactly one individual who controls all houses whenever $|\overline{I_s}| > 2$. We prove it in three steps (in accordance with Lemmas 1-3). Let s be the submatching created by the algorithm $\psi^{\kappa,\beta}$ before round r for $\succcurlyeq$.

- First, we show that an individual cannot own two houses while another individual owns a third house: By way of contradiction, suppose that some individual 1 owns house A and individual 2 owns houses B and C in round r. Then there exists an individual 3 who does not control any house at round r as $|\mathcal{H}| = |I|$. Consider four auxiliary preference profiles $\succcurlyeq^\ell$ that all share the following properties: (i) each individual matched under s ranks houses under $\succsim^{\ell}$, $\ell = 1, ..., 4$, in the same way they rank them under $\succsim$, (ii) each individual i unmatched at s and different from individuals $1, 2, 3$ ranks a unique s-unmatched house $H_i \notin \{A, B, C\} \cup \mathcal{H}_s$ as her first choice (such a unique house exists as $|\mathcal{H}| = |I|$), (iii) individuals 1 and 2 each rank all houses other than A, B, C lower than A, B, C, and (iv) individual 3's preference relation is the same as $\succsim_3$ under all four profiles. In particular, the four profiles differ only in how individuals 1 and 2 rank houses A, B, C: the ranking of A, B, C is the same as in the four preference profiles of the proof of Lemma 1 above. Notice that

$$\psi^{\kappa,\beta}[\succsim^{\ell}] = s \cup a^{\ell} \cup \{(i, H_i)\}_{i \in \overline{I_s} - \{1,2,3\}},$$

where a^{ℓ}s are defined as in the proof of Lemma 1 above. Furthermore, the same argument we used in the proof of Lemma 1 shows that there can be no SWF that ranks all four a^{ℓ}s, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{\kappa,\beta}$ efficient, a contradiction.

- Next, we show that one individual cannot control a house while at least two others each own a house in round r: Suppose, to the contrary, individual 1 owns house A, individual 2 owns house B, and individual 3 controls house C in round r. Consider three auxiliary preference profiles $\succsim^{\ell}$ that all share the following properties: (i) each individual matched under s ranks houses under $\succsim^{\ell}$, $\ell = 1, 2, 3$, in the same way they rank them under $\succsim$, (ii) each individual i unmatched at s and different from individuals $1, 2, 3$ ranks a unique s-unmatched house $H_i \notin \{A, B, C\} \cup \mathcal{H}_s$ as her first choice (such a unique house exists as $|\mathcal{H}| = |I|$), and (iii) individuals 1,2, 3 each rank all houses other than A, B, C lower than A, B, C, and the ranking of A, B, C is the same as in the three preference profiles of the proof of Lemma 2 above. Observe that
$$\psi^{\kappa,\beta}[\succsim^{\ell}] = s \cup a^{\ell} \cup \{(i, H_i)\}_{i \in \overline{I_s} - \{1,2,3\}},$$
where a^{ℓ}s are defined as in the proof of Lemma 2 above. Furthermore, the same argument we used in the proof of Lemma 2 shows that there can be no SWF that ranks all three a^{ℓ}s, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{\kappa,\beta}$ efficient, a contradiction.

- Finally, using a variant of Lemma 3, we show that there cannot be multiple brokers at round r (as multiple brokers can only occur with 3 individuals and 3 houses, where each individual brokers a distinct house): Suppose not. Then consider three auxiliary preference profiles $\succsim^{\ell}$ that all share the following properties: (i) each individual matched under s ranks houses under $\succsim^{\ell}$, $\ell = 1, 2, 3$, in the same way they rank them under $\succsim$, (ii) individuals 1,2,3, who are the only remaining unmatched individuals, each rank all houses other than A, B, C lower than A, B, C, and (iii) the ranking of A, B, C is the same as in the three preference profiles of the proof of Lemma 3 above. Notice that

$$\psi^{\kappa,\beta}[\succeq^{\ell}] = s \cup a^{\ell},$$

where a^{ℓ}s are defined as in the proof of Lemma 3 above. Furthermore, the same argument we used in the proof of Lemma 3 shows that there can be no SWF that ranks all three a^{ℓ}s, is transitive, and satisfies IIA. Hence, there is no complete SWF that makes $\psi^{\kappa,\beta}$ efficient, a contradiction.

Thus, a single individual owns all houses at round r when s is fixed for $|\overline{I_s}| > 2$ (by Corollary 4 and Remark 1).

This means that $\psi^{\kappa,\beta}$ is an almost sequential dictatorship, as all TC mechanisms restricted to only two individuals are almost sequential dictatorships.

($\Longleftarrow$) Consider an almost sequential dictatorship ψ^{κ}. If ψ^{κ} is a sequential dictatorship, then the proof of Theorem 4 works. So suppose it is not a sequential dictatorship. Hence, $|\mathcal{H}| = |I|$. We construct a complete SWF Φ such that ψ^{κ} is efficient with respect to Φ. Under Φ any two matchings are ranked according to the preference relation of the first-round dictator; if she is indifferent , then the matchings are ranked according to the preference relation of the second-round dictator, etc., until only two individuals remain unmatched. At this round let 1 and 2 be the two individuals and A and B be the two houses remaining unmatched. Observe that there are only two matchings, a and b, in which all individuals' assignments are the same but the last two: in one 1 gets A and 2 gets B, and in the other vice versa. Then one of these two matchings is equal to $\psi^{\kappa}[\succeq']$, where $\succeq'$ ranks the assignment of any individual other than 1 and 2 in a (or equivalently b) as her first choice, and for 1 and 2, the new preferences are the same as the original ones under $\succeq$. We rank $\psi^{\kappa}[\succeq'] \in \{a, b\}$ before the other one under $\Phi(\succeq)$.

Formally, for every $a \in \mathcal{A}$, let sequential dictators $i_1, \ldots, i_{|I|-2}$ be defined as $i_1 = \kappa_H(\emptyset)$ for every $H \in \mathcal{H}$, and in general, $i_{\ell} = \kappa_H(\{(i_1, a(i_1)), ..., (i_{\ell-1}, a(i_{\ell-1}))\})$ for every $H \in \mathcal{H} - \{a(i_1), ...a(i_{\ell-1})\}$ and $\ell = 1, ..., k$; then for every $b \in \mathcal{A} - \{a\}$, we say $a\, \Phi(\succeq)\, b$ if one of the following two conditions holds:

1. there exists $k \in \{1, ..., |I| - 2\}$ such that $a(i_1) = b(i_1)$, ..., $a(i_{k-1}) = b(i_{k-1})$, and $a(i_k) \succ_{i_k} b(i_k)$;

   or

2. for every $\ell \in \{1, ..., |I| - 2\}$, $a(i_{\ell}) = b(i_{\ell})$, and for $\succeq' \in \mathbf{P}$ where each i_{ℓ} ranks $a(i_{\ell})$ first while the remaining two individuals have the same preferences as in $\succeq$, we have $\psi^{\kappa}[\succeq'] = a$.

By construction, Φ is complete, antisymmetric, and transitive. Moreover, it satisfies the Pareto postulate. To see that it also satisfies IIA, consider two distinct matchings, $a, b \in \mathcal{A}$, and $\succeq \in \mathbf{P}$ such that $a\, \Phi(\succeq)\, b$. Also consider another profile $\hat{\succeq} \in \mathbf{P}$ such that each individual i's preference over the two matching assignments is the same in $\hat{\succcurlyeq}_i$ as in $\succcurlyeq_i$. If $a\ \Phi(\succcurlyeq)\ b$ because of condition 1 above, then condition 1 continues to hold for $\hat{\succcurlyeq}$ and thus $a\ \Phi(\hat{\succcurlyeq})\ b$. On the other hand, if $a\ \Phi(\succcurlyeq)\ b$ because of condition 2 above, then a and b only differ in how the last two individuals are assigned the remaining two houses. Hence, the profile constructed to check condition 2 for $a\ \Phi(\hat{\succcurlyeq})\ b$, which we refer to as $\hat{\succcurlyeq}'$, would lead to $\psi^{\kappa}[\hat{\succcurlyeq}'] = a$ because:

1. the first $|I| - 2$ dictators would still get their a assignments in the first $|I \vdash 2$ rounds of the TC algorithm for $\psi^{\kappa}[\hat{\succcurlyeq}']$, and

2. thus, the assignment of remaining two individuals under $\psi^{\kappa}[\hat{\succcurlyeq}']$ would be identical with that under a as the relative ranking of their assignments under a and b are identical both in $\succcurlyeq$ and $\hat{\succcurlyeq}$, and the ranking of the other houses do not matter for finding the outcome of the almost serial dictatorship.

Thus, $a\ \Phi(\hat{\succcurlyeq})\ b$, showing Φ satisfies IIA. QED

References

Abdulkadiroğlu, Atila and Tayfun Sönmez, "Random Serial Dictatorship and the Core from Random Endowments in House Allocation Problems," *Econometrica*, 1998, *66*, 689–701.

_ **and** _ , "House Allocation with Existing Tenants," *Journal of Economic Theory*, 1999, *88*, 233–260.

_ **and** _ , "School Choice: A Mechanism Design Approach," *American Economic Review*, 2003, *93*, 729–747.

Akbarpour, M and S Li, "Credible Auctions: A Trilemma," *Econometrica*, 2020, *88 (2)*, 425–467.

Arrow, Kenneth J., *Social Choice and Individual Values*, 2nd ed., New York: Wiley, 1963.

Austen-Smith, David and Jeffrey S. Banks, *Positive Political Theory I: Collective Preference*, University of Michigan Press, 1999.

Barberà, Salvador, "Strategy–Proofness and Pivotal Voters: A Direct Proof of the Gibbard–Satterthwaite Theorem," *International Economic Review*, 1983, *24*, 413—417.

_ **and Matthew O. Jackson**, "Strategy-proof Exchange," *Econometrica*, 1995, *63*, 51–87.

Bergson, Abram, "A Reformulation of Certain Aspects of Welfare Economics," *Quarterly Journal of Economics*, 1938, *52* (2), 310–334.

Black, Duncan, "On the Rationale of Group Decision-making," *Journal of Political Economy*, 1948, *56* (1), 23–34.

Bogolomania, Anna and Hervé Moulin, "Random Matching Under Dichotomous Preferences," *Econometrica*, 2004, *72*, 257–279.

Bogomolnaia, Anna and Hervé Moulin, "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 2001, *100*, 295–328.

Bordes, Georges and Michel Le Breton, "Arrovian theorems with private alternatives domains and selfish individuals," *Journal of Economic Theory*, 1989, *47* (2), 257–281.

_ **and** _ , "Arrovian theorems for economic domains: Assignments, matchings and pairings," *Social Choice and Welfare*, 1990, *7* (3), 193–208.

_ **and** _ , "Arrovian theorems for economic domains. The case where there are simultaneously private and public goods," *Social Choice and Welfare*, 1990, *7* (1), 1–17.

_ **, Donald E. Campbell, and Michel Le Breton**, "Arrow's Theorem for Economic Domains and Edgeworth Hyperboxes," *International Economic Review*, 1995, *36* (2), 441–454.

Campbell, Donald E. and Jerry S. Kelly, "A Strategy-Proofness Characterization of Majority Rule," *Economic Theory*, 2003, *22* (3), 557–568.

Curello, Gregorio and Ludvig Sinander, "Agenda-manipulation in Ranking," 2020. Working Paper.

Dasgupta, Partha, Peter Hammond, and Eric Maskin, "The Implementation of Social Choice Rules: Some General Results on Incentive Compatibility," *Review of Economic Studies*, 1979, *46*, 185–216.

Ehlers, Lars, "Coalitional Strategy-Proof House Allocation," *Journal of Economic Theory*, 2002, *105*, 298–317.

Ergin, Haluk, "Consistency in House Allocation Problems," *Journal of Mathematical Economics*, 2000, *34*, 77–97.

Gibbard, Allan, "Manipulation of Voting Schemes: A General Result," *Econometrica*, 1973, *41*, 587–601.

Hatfield, John W., "Strategy-Proof, Efficient, and Nonbossy Quota Allocations," *Social Choice and Welfare*, 2009, *33 No. 3*, 505–515.

Kalai, Ehud, Eitan Muller, and Mark A. Satterthwaite, "Social welfare functions when preferences are convex, strictly monotonic, and continuous," *Public Choice*, 1979, *34* (1), 87–97.

Klaus, Bettina and Eiichi Miyagawa, "Strategy-proofness, solidarity, and consistency for multiple assignment problems," *International Journal of Game Theory*, 2002, *30* (3), 421–435.

Le Breton, Michel and John A. Weymark, "Chapter Seventeen - Arrovian Social Choice Theory on Economic Domains," in Kenneth J. Arrow, Amartya Sen, and Kotaro Suzumura, eds., *Handbook of Social Choice and Welfare*, Vol. 2 of *Handbook of Social Choice and Welfare*, Elsevier, 2011, pp. 191–299.

Li, Shengwu, "Obviously Strategy-Proof Mechanisms," 2015. Working Paper.

Liu, Qingmin and Marek Pycia, "Ordinal Efficiency, Fairness, and Incentives in Large Markets," 2011. Working Paper.

Ma, Jinpeng, "Strategy-proofness and the strict core in a market with indivisibilities," *International Journal of Game Theory*, 1994, *23*, 75–83.

Maskin, Eric, "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 1999, *66*, 23–38.

Pai, Mallesh M. and Aaron Roth, "Privacy and Mechanism Design," 2018. Working Paper.

Pápai, Szilvia, "Strategyproof Assignment by Hierarchical Exchange," *Econometrica*, 2000, *68*, 1403–1433.

_ , "Strategyproof and Nonbossy Multiple Assignments," *Journal of Public Economic Theory*, 2001, *3*, 257–271.

_ , "Exchange in a General Market with Indivisible Goods," *Journal of Economic Theory*, 2007, *132*, 208–235.

Pycia, Marek, "Swaps on Networks," 2016. Working Paper.

_ , "Evaluating with Statistics: Which Outcome Measures Differentiate Among Matching Mechanisms?," 2019. Working paper.

_ **and M. Utku Ünver**, "Outside Options in Neutral Discrete Resource Allocation," 2007. Working paper.

_ **and** _ , "Incentive Compatible Allocation and Exchange of Discrete Resources," *Theoretical Economics*, 2017, *12*, 287–329.

_ **and Peter Troyan**, "A Theory of Simplicity on Games and Mechanism Design," 2019. Working paper.

Root, Joseph and David Ahn, "Incentives and Efficiency in Constrained Allocation Mechanisms," 2020. Working Paper.

Roth, Alvin E., Tayfun Sönmez, and M. Utku Ünver, "Kidney Exchange," *Quarterly Journal of Economics*, 2004, *119*, 457–488.

Samuelson, Paul A., *Foundations of Economic Analysis*, MA: Harvard University Press, 1947.

Satterthwaite, Mark, "Strategy-proofness and Arrow's Conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 1975, *10*, 187–216.

_ **and Hugo Sonnenschein**, "Strategy-Proof Allocation Mechanisms at Differentiable Points," *Review of Economic Studies*, 1981, *48*, 587–597.

Sen, Amartya, "Interpersonal Aggregation and Partial Comparability," *Econometrica*, 1970, *38* (3), 393–409.

_ , "The Possibility of Social Choice," *American Economic Review*, 1999, *89* (3), 349–378.

Svensson, Lars-Gunnar, "Queue Allocation of Indivisible Goods," *Social Choice and Welfare*, 1994, *11*, 223–230.

_ , "Strategy-proof Allocation of Indivisible Goods," *Social Choice and Welfare*, 1999, *16*, 557–567.

Takamiya, Koji, "Coalition Strategy-Proofness and Monotonicity in Shapley-Scarf Housing Markets," *Mathematical Social Sciences*, 2001, *41*, 201–213.

Tang, Qianfeng and Yongchao Zhang, "Hierarchical Exchange Rules and the Core in Indivisible Objects Allocation," 2015.

Weymark, John, "Arrow's theorem with social quasi-orderings," *Public Choice*, 1984, *42*, 235–246.

Woodward, Kyle, "Self-Auditable Auctions," 2020. Working paper.

Zhou, Lin, "Inefficiency of Strategy-Proof Allocation Mechanisms in Pure Exchange Economies," *Social Choice and Welfare*, 1991, *8*, 247–257.

Zwicker, William S., *Introduction to the Theory of Voting*, Cambridge University Press,

A Omitted Proof

Proof of Theorem 3. (Group strategy-proofness $\implies$ individual strategy-proofness and non-bossiness) By definition, any group strategy-proof mechanism is immune to all single-person group deviations, and hence, it is also individually strategy-proof. To the contrary to the claim, suppose a group strategy-proof mechanism φ is not non-bossy. Then there exists some individual i, preference profile $\succcurlyeq$, and i's preference relation $\succcurlyeq_i'$ such that $a = \varphi[\succcurlyeq] \equiv_i \varphi[\succcurlyeq_i', \succcurlyeq_{-i}] = a'$ and yet there exists some individual $j \neq i$ such that $a \not\equiv_j a'$. Consider the group $J = \{i, j\}$. By richness assumption, we have either $a \succ_j a'$ or $a' \succ_j a$. If the former is the case, then consider the group deviation $(\succcurlyeq_i, \succcurlyeq_j)$ from the profile $(\succcurlyeq_i', \succcurlyeq_{-i})$: individual i is indifferent while individual j is better off contradicting group strategy-proofness of φ. If the latter is the case, then consider the group deviation $(\succcurlyeq_i', \succcurlyeq_j)$ from the profile $(\succcurlyeq_i, \succcurlyeq_{-i})$. Individual i is indifferent while individual j is better off, again contradicting group strategy-proofness of φ. Thus, we showed that φ is also non-bossy.

(Individual strategy-proofness and non-bossiness $\implies$ monotonicity) Let φ be an individual strategy-proof and non-bossy mechanism. Consider a preference profile $\succcurlyeq$. Let $\succcurlyeq' \in \mathbf{P}$ be one of its φ-monotonic transformations. We prove this part by induction. Suppose as the inductive assumption, we proved that for a given $J \subset I$ (for the base case $J = \emptyset$ trivially holds), we showed that $\varphi[\succcurlyeq_J', \succcurlyeq_{-J}] \equiv_j \varphi[\succcurlyeq_{-J}]$ for every $j \in I$. Consider an individual $i \in I - J$. Let $\tilde{J} = \{i\} \cup J$. First we establish that $\varphi[\succcurlyeq_{\tilde{J}}', \succcurlyeq_{-\tilde{J}}] \equiv_i \varphi[\succcurlyeq]$: Suppose not, to the contrary of the claim. Let $a' = \varphi[\succcurlyeq_{\tilde{J}}', \succcurlyeq_{-\tilde{J}}] \not\equiv_i \varphi[\succcurlyeq_J', \succcurlyeq_{-J}] = a \equiv_i \varphi[\succcurlyeq]$. If $a' \succ_i' a$, then $a \succ_i' a'$ by construction of $\succcurlyeq_i'$ and this contradicts individual strategy-proofness of φ for i, as she can report $\succcurlyeq_i'$ and be better off while her preference relation is $\succcurlyeq_i$ and others have preferences $(\succcurlyeq_J', \succcurlyeq_{-\tilde{J}})$. If $a \succ_i' a'$, this contradicts individual strategy-proofness of φ for i, as she can report $\succcurlyeq_i$ and be better off while her preference relation is $\succcurlyeq_i'$ and others have preferences $(\succcurlyeq_J', \succcurlyeq_{-\tilde{J}})$. Thus, $a \sim_i a'$. Since $a \not\equiv_i a'$, this last statement contradicts part 2 of the richness assumption. Thus $a \equiv_i a'$. Then non-bossiness of φ implies that $a \equiv_j a'$ for every $j \in I$. Inductive assumption implies that $\varphi[\succcurlyeq_{\tilde{J}}', \succcurlyeq_{-\tilde{J}}] \equiv_j \varphi[\succcurlyeq_{-\tilde{J}}]$ for every $j \in I$.

(Monotonicity $\Longrightarrow$ group-strategy-proofness). Let φ be a monotonic mechanism. Consider a preference profile $\succcurlyeq$, a group $J \subseteq I$, and a possible deviation $\succcurlyeq'_J$. Suppose $a' = \varphi[\succcurlyeq'_J, \succcurlyeq_{-J}] \succcurlyeq_j \varphi[\succcurlyeq] = a$ for every $j \in J$ and for some individual $i \in J$ the preference relation is strict. Consider the preference profile of J, $\succcurlyeq^*_J$ such that a' is ranked higher than a and every other equivalence class of alternatives are ranked below these two alternatives' equivalence classes. $(\succcurlyeq^*_J, \succcurlyeq_{-J})$ is a φ-monotonic transformation of $\succcurlyeq$, and hence, $\varphi[\succcurlyeq^*_J, \succcurlyeq_{-J}] \equiv_j a$ for al $j \in I$ by monotonicity of φ. Since a' is the top alternative in $\succcurlyeq^*_j$ for every $j \in J$ and $\varphi[\succcurlyeq'_J, \succcurlyeq_{-J}] = a'$, $(\succcurlyeq^*_J, \succcurlyeq_{-J})$ is also φ-monotonic transformation of $(\succcurlyeq'_J, \succcurlyeq_{-J})$, and hence, $\varphi[\succcurlyeq^*_J, \succcurlyeq_{-J}] \equiv_j a'$ for every $j \in I$ by monotonicity of φ. Since $a \not\equiv_i a'$, we obtain a contradiction. Thus, φ is group strategy-proof. QED

B An Incomplete Arrovian Social Welfare Function

The following example illustrates an incomplete Arrovian SWF.

Example 3: Consider a society (or an employer) assigning one task to each of three employees. All the tasks need to be completed, and the society would like to respect the preferences of the employees in assigning the tasks as much as possible. As a second order concern, the society would like to avoid assigning Task A to employee 1 (e.g. because of a belief that employee 1 is not very good in doing this job). The society thus has an SWF that has the maximum at a Pareto-efficient matching that does not assign Task A to employee 1 if there exists at least one Pareto-efficient matching that does not assign Task A to employee 1.

The society's SWF can be equivalently described in terms of a Trading Cycles mechanism ψ in which employee 1 brokers A, employee 2 has ownership of B and employee 3 has ownership of C: for any preference profile $\succcurlyeq_{\{1,2,3\}}$, the SWF $\Psi(\succcurlyeq)$ ranks any two distinct matchings a and b if and only if $a = \psi[\succcurlyeq]$ or a Pareto dominates b; the social ranking is then $a \; \Psi(\succcurlyeq) \; b$.

For instance, for the preference profile

$$\succcurlyeq = \begin{array}{|c|c|c|} 1 & 2 & 3 \\ \hline A & A & B \\ B & B & C \\ C & C & A \end{array},$$

the outcome of Trading Cycles ψ is $\psi[\succcurlyeq] = \{(1,B),(2,A),(3,C)\}$, and the ranking of the matchings with respect to $\Psi(\succcurlyeq)$ is given in Figure 1.

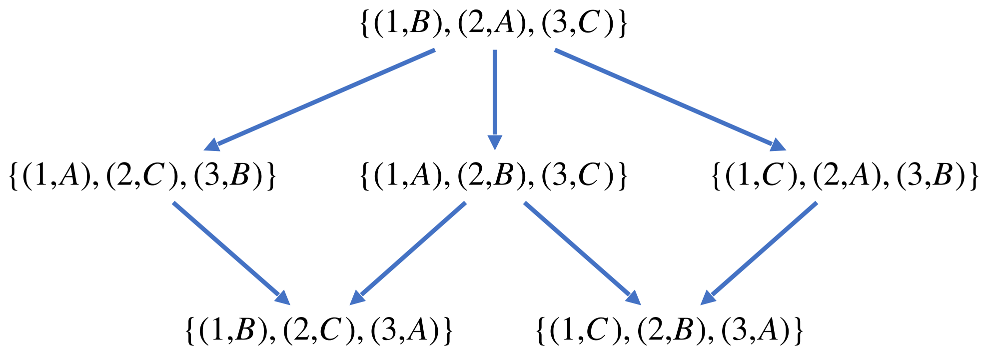

Figure 1: $\Psi(\succcurlyeq)$ in Example 3. For matching a, b, we have $a\ \Psi(\succcurlyeq)\ b$ if and only if there is a directed path from a to b in this graph.

$\blacksquare$