A Theory of Auditability for Allocation and Social Choice Mechanisms*

Aram Grigoryan† Markus Möller‡

August 2023

Abstract

In centralized mechanisms and platforms, participants do not fully observe each others' type reports. Hence, the central authority (the designer) may deviate from the promised mechanism without the participants being able to detect these deviations. In this paper, we develop a theory of auditability for allocation and social choice problems. Namely, we measure a mechanism's auditabilty as the size of information needed to detect deviations from it. We apply our theory to study auditability in a variety of settings. For priority-based allocation and auction problems, we reveal stark contrasts between the auditabilities of some prominent mechanisms. For house allocation problems, we characterize a class of dictatorial rules that are maximally auditable. For voting problems, we provide simple characterizations of dictatorial and majority voting rules through auditability. Finally, for choice with affirmative action problems, we study the auditability properties of different reserves rules.

*We thank seminar participants at Bonn, Boston College, Caltech, Duke, UC Berkeley Duke, UC San Diego, and conference participants at the Iowa State University Market Design workshop, Stony Brook Game Theory conference, Southwest Economic Theory conference, Conference on Economic Design, ACM FAccT'23, and EC'23 for helpful discussions and comments. This paper is based on an earlier paper by Aram Grigoryan, titled 'Transparency in Allocation Problems' (March 2022).

†Department of Economics, University of California - San Diego, ag@ucsd.edu

‡Department of Economics, University of Bonn, mmoelle2@uni-bonn.de

1 Introduction

Scarce resources are oftentimes allocated in a centralized clearinghouse; the designer collects participants' type reports, and chooses an allocation through some publicly announced mechanism. Some major examples include allocation of public school seats or subsidized housing, auctions for online ads or spectrum licences, and presidential elections. In these problems, participants know and observe their own type reports and outcomes, but not necessarily those of others. Hence, if the designer uses a mechanism that is different from the promised one, participants or an auditing entity may not be able detect this deviation because of the limited information. In this paper, we develop a general theory of auditability to compare mechanisms in terms of how easy it is for participants or some third-party auditing entity to detect such deviations.

We apply our theory across a wide range of applications such as allocation problems, single-object auctions, voting with a binary outcome, and choice with affirmative action. In all these settings, auditability can be an important consideration given the practical concerns about errors, misconduct and fraud during the implementation stage. For example, there have been multiple controversies during the implementation of admissions to Chicago Public Schools (CPS).[1] The Office of Inspector-General (OIG) of Chicago Board of Education (Schuler, 2018) states that in the 2016-2017 admission year:

> *"...almost every kind of CPS elementary school imaginable improperly admitted students last school year. ... Of more than 18,200 elementary-grade admissions audited, nearly 6,900 failed the audit. That's nearly two of every five."*

The report further states that:

[1] More generally, the need for transparency and auditability in public school assignment has been highlighted in Pathak (2017) and Benner and Boser (2018). The latter argues that the assignment mechanism should be audited by an outside entity to guarantee that the implementation is consistent with enrollment priorities. Such auditing practices have been adopted by school districts in New Orleans and Chicago (Benner and Boser, 2018).

"OIG interviews with principals of 30 audited schools that held more than 500 combined audit failures revealed that many didn't know all the admissions rules, which are scattered across several locations. Others knew the rules and broke them. In some cases, audit failures may have been caused by documentation errors. ... Several principals weeded out applicants, based on a variety of factors, including attendance concerns. ... Some principals clearly played favorites. Many schools bypassed [admission rules] to give preference to the children of CPS employees, the siblings of existing students or lottery winners, ... One principal improperly admitted her four children, her niece and nephew."

Auditability issues are also salient in online ads auctions, which is largely driven by the lack of bid-observability in these markets. Recently, there have been various allegations against Google's online ads auctions related to issues of auditability. In 2020, an antitrust lawsuit brought by a coalition of states led by Texas attorney general Ken Paxton accused Google of not implementing the auctions in the way they promised. The Wired magazine writes:[2]

"... the most recent complaint [against Google] ... provides unprecedented insight into how Google allegedly misled advertisers and publishers for years by manipulating auctions in its own favor using inside information. As one employee put it in a newly revealed internal document, Google's public claim about second-price auctions were 'untruthful.'"

In fact, theoretical concerns about the scope for auctioneers to manipulate outcomes in the second-price auction are not new (Vickrey, 1961), and they have received some attention at different points in the media. For example, as the Digiday newsletter writes:[3]

"In a second-price auction, raising the price floors after the bids come in allows [online auctioneers] to make extra cash off unsuspecting buyers [...] This practice persists because neither the publisher nor the ad buyer has complete

[2] https://www.wired.com/story/google-antitrust-ad-market-lawsuit/

[3] https://digiday.com/marketing/ssps-use-deceptive-price-floors-squeeze-ad-buyers/

> *access to all the data involved in the transaction, so unless they get together and compare their data, publishers and buyers won't know for sure who their vendor is ripping off ...*"

Motivated by the salience of auditability considerations in different environments, we study auditability in a general model that covers a variety of settings. In our model, there are N individuals and a commonly known mechanism which specifies an outcome for each realization of individuals' type reports. Given a profile of type reports, which we call a problem, a deviation from a mechanism describes a situation where the implemented outcome is different from the one promised by the mechanism. For a given problem and a deviation, we say that a group of individuals detects the deviation from the mechanism, if the group's outcome could not have been resulted by the mechanism for any problem that is consistent with the group members' type reports. We then define two straightforward and index-based measures that allow us to compare mechanisms' auditability properties on different dimensions. First, for a given problem, the **auditability index** of a mechanism is the smallest integer such that for any deviation, there is a group of individuals whose number does not exceed this integer and who detects the deviation. Second, we define the **worst-case auditability index** of a mechanism as its largest auditability index across all problems. The main idea is that mechanisms with small auditability indices may be easier to audit, while those with large auditability indices are more vulnerable to undetected errors or fraud.

In our main application, we study some well-known and practically applied mechanisms for allocation without money and unveil sharp contrasts in their auditability properties. For general object allocation problems, we give a simple characterization of the entire class of mechanisms with auditability index of one (Theorem 1). We establish that such mechanisms are highly restrictive, in a sense that they violate very basic notions of efficiency (Proposition 1). Hence, the reasonable range for the auditability index of object allocation mechanisms falls between 2 and N.

Priority-Based Allocation. In the context of priority-based allocation, we investigate the auditability properties of a large class of mechanisms that are outcome-equivalent to the Deferred Acceptance mechanism of Gale and Shapley (1962) applied to systematically modified problems. We call them *Deferred-Acceptance (DA)-representable* mechanisms. Among many other well-studied mechanisms, this class covers the entire family of *application-rejection* rules (Chen and Kesten, 2017), which includes the prominent Deferred Acceptance and Immediate Acceptance (aka the 'Boston' mechanism) as special cases.

For the general class of DA-representable mechanisms, we provide a tight characterization result (that is, necessary and sufficient conditions) for having a worst-case auditability index of two (Theorem 2). The characterization reveals close connection between a mechanism's low auditability index and the structure of *stable* outcomes in the DA-representation.

As a corollary of our characterization result, we establish that Immediate Acceptance has a worst-case auditability of two (Proposition 1). Interestingly, we find that the Immediate Acceptance is the unique mechanism in the application-rejection family that has good auditability properties: all other mechanisms in the family, including Deferred Acceptance, have a worst-case auditability index of N (Proposition 4).

Our findings may contribute to the ongoing discussions and evaluations of popular mechanisms for practical application. Traditionally, many school districts in the US have been using Immediate Acceptance for centralized assignment. Starting from the early 2000s, many major school districts transitioned to the Deferred Acceptance mechanism, potentially because the latter mechanism is *strategy-proof*. That means, in theory, parents do not have to worry about finding the best strategy for ranking schools (Abdulkadiroğlu and Sönmez, 2003). Currently, Deferred Acceptance is used for student assignment in Boston, New York City, Denver, Chicago, New Orleans, Newark, and Indianapolis. Despite these transitions, the Immediate Acceptance still remains one of the most common school assignment mechanisms. The mechanism is used in Charlotte-Mecklenburg, Miami-Dade, Minneapolis, Seattle, Tampa-St. Petersburg, and many other school districts in the US and around the world. To the best of our knowledge, we provide a first theoretical comparison of Immediate Acceptance and Deferred Acceptance in terms of a new interesting and important dimension. More specifically, we find that Immediate Acceptance has superior performance in terms of auditability. The findings potentially highlight the importance of addressing trust and transparency considerations, in order to maximize the adoption success of the strategically simple and superior Deferred Acceptance for real-life assignment problems.

House Allocation. In the context of house allocation (Hylland and Zeckhauser, 1979) we focus on the large class of *sequential dictatorships*.[4] It is well-known that these mechanisms satisfy various desirable properties such as Pareto efficiency, non-bossiness (Satterthwaite and Sonnenschein, 1981) and group strategy-proofness. In addition, Pycia and Troyan (2023) showed more recently that sequential dictatorships are among the few candidates that achieve high standards for strategic simplicity.[5] Our main result (Theorem 3) in this setting is a full characterization of maximally auditable sequential dictatorships through a new class of mechanisms of which all members are *almost* serial dictatorial. We also establish that the range of worst-case indices for sequential dictatorships is wide and that even members that are close to a serial dictatorship can be hard to audit (Proposition 5). Thus, our results suggest that one cannot go much beyond simple dictatorial orderings if one wants to assure that the mechanism remains easy to audit in our sense.

Single-Object Auction. We also demonstrate significant differences in the auditabilities of some prominent single-object auction mechanisms. As a starting point, we establish that auctions with a worst-case auditability index of one are characterized by a very restrictive form of dictatorship: there must exist two individuals of which one individual receives the object, whenever she reports a bid from a fixed subset of her bids. However, whenever she reports a bid not in that set, the object is assigned to the other individual (Theorem 4). In light of this result, it turns out that two popular auction formats can be regarded as close to being maximally auditable. Both the standard first-price auction mechanism and the all-pay auction mechanism have

[4] In a sequential dictatorship (Ehlers and Klaus, 2003; Papai, 2000; Pápai, 2001) individuals sequentially choose their favorite objects from those that are still available. Who is the next individual in the ordering may depend on all previous individuals' assignments and hence its determination can be complex. When individuals choose according to a fixed ordering, then a sequential dictatorship reduces to a *serial dictatorship* (Satterthwaite and Sonnenschein, 1981).

[5] Pycia and Troyan (2023) show that sequential dictatorships closely resemble the entire class of mechanisms that are *strongly obviously strategy-proof* for the context of object allocation without money.

a worst-case auditability index of exactly two (Proposition 6). On the other extreme, we find that the second-price auction mechanism is minimally auditable by showing that it has an auditability index of N for *any* profile of bids (Proposition 7).

Voting with a Binary Outcome. For voting problems, we first show that a worst-case auditibility index of one requires the social choice mechanism to be dictatorial (Theorem 5). We then restrict attention to the class of *anonymous* social choice mechanisms for which all individuals are treated symmetrically. Interestingly, our theory yields a simple and new characterization of the *majority voting* mechanism as the unique most auditable anonymous mechanism. We show that the majority voting is the unique mechanism with the smallest worst-case auditability index among anonymous ones (Theorem 6).

Choice with Affirmative Action. Our theory can also address an important design question about reserves processing orders in choice with affirmative action and distributional objectives. When choosing a subset of individuals for a limited number of positions, a reserves mechanism first sets aside (or reserves) several positions for a particular beneficiary group, and prioritizes the group for these (reserved) positions. Reserves mechanisms have been used in controlled school choice, (Hafalir, Yenmez, and Yildirim, 2013), college admissions in Brazil (Aygun and Bó, 2021), allocation of visa lotteries (Pathak, Rees-Jones, and Sönmez, 2020), pandemic rationing (Pathak, Sönmez, Ünver, and Yenmez, 2021), and affirmative action in India (Sönmez and Yenmez, 2022), among many other applications.

The order in which applicants are considered for reserved versus non-reserved (open) seats affects the final outcome (Dur, Kominers, Pathak, and Sönmez, 2018). The *reserved-seats-first* and the *open-seats-first* are two well-studied implementations of reserves mechanisms. Given the non-trivial implications of reserves processing orders, researchers and policy makers may be interested in the choice of the 'best' reserves mechanism. Some recent papers provide an axiomatic foundation for the reserved-seats-first mechanism, where the axioms reflect the policy objectives of meeting distributional objectives and respecting priorities (Abdulkadiroğlu and Grigoryan, 2021; Echenique and Yenmez, 2015). Our auditability theory provides a different angle for comparing reserves mechanisms. We find that the reserved-seats-first mechanism is more auditable than the open-seats-first mechanism (Proposition 8)

The remainder of this work is organized as follows. In section 2, we discuss the related literature. Section 3 introduces the model and our notions of auditability. Section 4 applies our theory to object allocation without money, section 5.1 to single-object auctions, section 5.2 to voting, and section 5.3 to choice with affirmative action. Section 6 discusses potential extensions and provides further considerations. Section 7 concludes. Proofs not in the main text are in Appendix A.

2 Related Literature

Our paper contributes to the relatively recent strand of literature on auditability, transparency, and credibility of mechanisms. For auction mechanisms, these notions have been studied by Akbarpour and Li (2020) and Woodward (2020), and for matching mechanisms, by recent and concurrent works of Hakimov and Raghavan (2020), Pycia and Ünver (2022), and Möller (2022). The existing literature typically studies specific environments and a binary classification of mechanisms - (fully) auditable versus not auditable - where an auditable mechanism is one where deviations can always be noticed by a single individual. Full auditability is hard to achieve in the static environment and without additional communication, and therefore most papers on the topic include dynamic elements with ongoing private or public communication from the designer. We do not explicitly model any form of communication. Instead, we quantify a mechanism's auditability with a natural informational measure—the size of the group whose information is sufficient for detecting deviations from the mechanism. Hence, whereas other papers on the topic study specific environments and communication protocols, we introduce a general theory for auditability that is applicable (and informative) for a variety of allocation and social choice problems.

Although our framework and theory substantially differ from the previous and concurrent works on the topic, some of our findings about specific mechanisms are related to certain observations from these other works. For example, Akbarpour and Li (2020) show that the first-price auction is credible, whereas the second-price auction is not credible. We find a stark contrast between the auditabilities of these mechansims using our non-binary comparison measure. Hakimov and Raghavan (2020) observe that in general the DA is not 'transparent'. Möller (2022) shows that when school-specific priorities are publicly known, DA achieves transparency if and only if it is a serial dictatorship.[6] Unlike the other papers, we do not explicitly model any communication or public knowledge. Instead, we introduce a (non-binary) informational measure which allows to quantify the degree of (non-)auditability of DA. We establish that the worst-case auditability index of the DA is maximally large. Moreover, we fully characterize all problems where the mechanism's auditability index is exactly two.

On a higher level, our study of auditability contributes to the theoretical understanding about the structure of matching mechanisms. Naturally, the notion of auditability is closely related to the question of how a single individual (or a group of individuals) can affect the outcome of the mechanism by only changing her own type report. This question has been discussed by Arnosti (2020), who note that the DA mechanism is 'unpredictable' in this sense. Gonczarowski and Thomas (2023) also explore this problem, and also study a parallel notion of 'verifiability' in matching problems. Gangam, Mai, Raju, and Vazirani (2023) analyze the notion of 'robust' stable matching, which is simultaneously a stable matching for two problems that only differ by one type report. Interestingly, the authors highlight implications of their theory to some notion of credibility or auditability for stable mechanisms.

Lastly, our paper relates to the literature on the role of privacy in market design. Without privacy concerns, the central authority could achieve more transparency by some (credible) communication of all participants' type reports and outcomes. When there are privacy concerns, one would prefer that transparency is achieved with minimal informational leakage. One implication of our auditability index is that it measures the smallest amount of private information needed for a successful audit. Hence, more auditable mechanisms achieve trust and transparency with less privacy costs.

[6] Möller (2022) also characterizes the entire class of transparent, efficient and strategy-proof mechanisms as the class of sequential dictatorships. Crucially, our results on sequential dictatorships are not logically connected to those in Möller (2022).

There is a line of research in the auctions literature that studies the possibilities of improving auditability and transparency with cryptographic protocols (e.g., Brandt (2001); Ferreira and Weinberg (2020)) and blockchain (e.g. (Chitra, Ferreira, and Kulkarni, 2023)). In a more general environment, Canetti, Fiat, and Gonczarowski (2023)) study the possibility of trustworthy mechanisms that preserve the privacy of the central authority (i.e., reveal no information about the mechanism), as opposed to preserving privacy of market participants. To our knowledge, the privacy question in object allocation (without money) is relatively underexplored. In a recent paper, Haupt and Hitzig (2023) examine privacy-preserving implementation (without any cryptographic commitments) of prominent object allocation mechanisms and auctions through 'sequential elicitation protocols'. Specifically, whereas our auditability notion tells how much privacy can be preserved under a successful audit, Haupt and Hitzig (2023) ask whether privacy can be protected in the sense of not eliciting information that is inconsequential for computing the final outcome. Finally, Ollar, Rostek, and Yoon (2021) analyze the influence of privacy on efficiency in dynamic markets from a market design perspective.[7]

3 Mechanisms and Auditability

There is a set of individuals $\mathcal{I}$, a set of options Ω_i for each $i \in \mathcal{I}$, and a set of feasible outcomes $\Omega \subseteq \times_{i \in \mathcal{I}} \Omega_i$. Each individual i has a type report $\theta_i \in \Theta_i$, and $\Theta \subseteq \times_{i \in \mathcal{I}} \Theta_i$ denotes the set of feasible type report profiles. We refer to an element $\theta := (\theta_i)_{i \in \mathcal{I}} \in \Theta$ as a *problem*. A mechanism is a mapping $\varphi : \Theta \to \Omega$ that gives a feasible outcome $\varphi(\theta)$ for every problem $\theta \in \Theta$.

For a problem $\theta \in \Theta$, and a subset of individuals $I \subseteq \mathcal{I}$, define $\theta_I := (\theta_i)_{i \in I} \in \times_{i \in I} \Theta_i := \Theta_I$ and $\theta_{-I} := (\theta_i)_{i \in \mathcal{I} \setminus I} \in \times_{i \in \mathcal{I} \setminus I} \Theta_i := \Theta_{-I}$.

We measure the auditability of a mechanism by how easy it is for individuals to detect deviations from it. Formally, a deviation from mechanism φ at problem θ is

[7] For a recent survey on privacy in economic theory see, for instance, Acquisti, Taylor, and Wagman (2016).

an outcome $\omega \neq \varphi(\theta)$. We say that the non-empty set of individuals $I \subseteq \mathcal{I}$ **detects the deviation** $\omega \neq \varphi(\theta)$ if for any $\theta'_{-I} \in \Theta_{-I}$,

$$\omega(i) \neq \varphi(\theta_I, \theta'_{-I})(i) \text{ for some } i \in I.$$

In other words, a deviation is detected by a group of individuals if the group's information, consisting of their type reports and assignments, is sufficient to rule out that the promised mechanism was used.

We define the auditability index of a mechanism as the smallest integer such that for any deviation, there is a subset of individuals whose cardinality does not exceed that integer and who detect the deviation.

Definition 1. *Consider a mechanism φ.*

(*i*) *The **auditability index** of φ for a given problem θ is*

$$\#\varphi^\theta = \max_{\omega \neq \varphi(\theta)} \min \Big\{ |I| : I \subseteq \mathcal{I}, I \text{ detects the deviation} \Big\}.$$

(*ii*) *The **worst-case auditability index** of φ is*

$$\#\varphi = \max_{\theta \in \Theta,\, \omega \neq \varphi(\theta)} \min \Big\{ |I| : I \subseteq \mathcal{I}, I \text{ detects the deviation} \Big\}.$$

The basic idea behind the definitions is the following: a lower auditability index means that less information is needed to detect deviations. Hence, mechanisms with a lower auditability index are easier to audit.

We do not provide any explicit microfoundations for how information is collected or accessed. However, we believe that our auditability index is a natural measure for information, and one can find many examples and stories where such a measure is informative. For instance, one could imagine that some individuals observe some other individuals' type reports, or individuals share reports with each other, or a third party auditing entity samples a subset of type reports and outcomes.

Note that if we treat each individual's private information as an atom—a smallest piece of information—then our auditability index can be interpreted as the number of pieces of information needed for detecting deviations. Hence, our auditability index may be deemed as a natural benchmark for comparing auditability. More generally, we believe that our theory will also be informative for alternative notions of auditability, such as a probabilistic one. We elaborate on this point in Section 6.

4 Auditability in Object Allocation

4.1 Preliminaries

Consider the problem of allocating a set of objects $\mathcal{O}$ to a set of individuals $\mathcal{I}$, with $|\mathcal{I}| = |\mathcal{O}| = N \geq 2$. The setup corresponds to a special case of our model where $\Omega_i \equiv \mathcal{O}$ for all $i \in \mathcal{I}$, and $\Omega \subseteq \mathcal{O}^{\mathcal{I}}$ is the space of one-to-one functions $\omega : \mathcal{I} \to \mathcal{O}$.

For now, we will keep the setup general and will not specify what a type report θ_i of individual i stands for. In next subsections, we will study two special cases separately: In the first model the type reports represent preference rankings and priority scores over objects (priority-based allocation), and in the second model the type reports will be preference rankings over objects only (house allocation).

We first provide a tight characterization of the entire class of object allocation mechanisms with (worst-case) auditability index of one. The characterization will pin down a very restrictive class of mechanisms, that will fail to satisfy mild (efficiency) properties.

Fix a mechanism φ. We say object o is **possible** for individual i at type report θ_i if $o = \varphi(\theta_i, \theta_{-i})$ for some θ_{-i}. We say that an individual i **clinches** object o at type report θ_i, if o is the only possible object for i when she reports θ_i. Given a subset of objects $O \subseteq \mathcal{O}$, we say that i clinches o from set O at type report θ_i, if o is the only possible object among O for i at θ_i. That is, there is no $o' \in O \setminus \{o\}$ such that $\varphi(\theta_i, \theta_{-i})(i) = o'$ for some θ_{-i}.

Definition 2. *We say φ has a **sequential clinching implementation** at problem θ, if $\varphi(\theta)$ is an outcome of the following algorithm:*

Step 0*. Initially all individuals and objects are available. Start with Step* 1*.*

Step t*. Pick an available individual* i *who clinches some object* o *among the available ones when she reports* θ_i*.*[8] *Make* i *and* o *unavailable. If there is no available individual, the algorithm terminates. Otherwise, proceed to Step* $t+1$*.*

We characterize mechanisms with sequential clinching implementation as the unique mechanisms with an auditability index of one.

Theorem 1. *Consider an arbitrary mechanism* φ*.*

1. *For any* $\theta \in \Theta$*,* $\#\varphi^\theta = 1$ *if and only if* φ *has a sequential clinching implementation at* θ*.*

2. $\#\varphi = 1$ *if and only if* φ *has a sequential clinching implementation at any problem* θ*, and moreover, the mechanism can be implemented with a sequential clinching order that only depends on the set of available objects at each step (but otherwise does not depend on the problem).*

Proof. See Appendix A.1. □

Being able to clinch an object is a strong restriction on the mechanism.[9] To emphasize the limitations of sequential clinching implementation, we demonstrate that all these mechanisms violate basic efficiency notions when type reports represent standard preferences (as in the next sections' applications). In particular, we show that mechanisms with an auditability index of one do not satisfy the full range property. We say a mechanism ϕ has **full range** if for any outcome $\omega \in \Omega$, there is a problem

[8] We implicitly assume that such an individual exists, which is a strong restriction and a defining feature of the sequential clinching implementation.

[9] Sequential clinching implementation is substantially different from serial dictatorial implementation which we study in Section 4.3. The former is much more restrictive. In fact, for a 'generic' problem serial dictatorships do not have a sequential clinching implementation. We discuss these points further in Appendix C.

θ such that $\varphi(\theta)(i) = \omega(i)$ for all $i \in \mathcal{I}$.[10] In other words, a mechanism has full range if every outcome may be selected for some problem.

Proposition 1. *If $\#\varphi = 1$, then φ does not have a full range.*

Proof. See Appendix A.2. $\square$

If individuals' type reports represent their preferences over objects, then the full range property is implied by a very weak notion of efficiency which says that a mechanism should always assign everyone to their most preferred objects whenever it is feasible to do so. Hence, almost any reasonable mechanism will have a worst-case auditability index of weakly larger than two. Interestingly, as we will see in the next subsections, many highly applicable mechanisms achieve this lower bound, that is, they have a worst-case auditability index of exactly two.

4.2 Priority-Based Allocation

The priority-based allocation model covers various real-life assignment problems, including school choice and college admissions. The type report of an individual $i \in \mathcal{I}$ is $\theta_i = (P_i, r_i)$, where P_i is a strict preference ranking over objects $\mathcal{O}$, and $r_i \in \mathbb{R}^N$ is a vector of object-specific priority scores for individual i. For individuals i and j, and an object $o \in \mathcal{O}$, we say that i has a higher priority at o than j if $r_{io} > r_{jo}$. Let $P = (P_i)_{i \in \mathcal{I}}$ and $r = (r_i)_{i \in \mathcal{I}}$ denote the profile of preferences and priorities, respectively. We assume strict priorities, that is, for any $\theta = (P, r) \in \Theta$, any $i, j \in \mathcal{I}$, and $o \in \mathcal{O}$, we assume that $r_{io} \neq r_{jo}$.[11]

[10] In the context of house allocation (Section 4.3), the condition is also known as *citizen sovereignty* (Pápai, 2001).

[11] In some allocation problems, such as school choice, applicants are categorized into a few priority classes, and strict priorities are obtained only after ties are broken using randomly drawn lottery numbers. In our setup, the r's denote this strict priority scores obtained after the tie-breaking, i.e., the priority group plus the random number. We think of the refined priorities as the private information in our model. In several school districts, such as the New York City, the random numbers are privately known by applicants, which is consistent with our setup.

Given a problem $\theta = (P, r)$, we say that o is the n-th most preferred object for individual i at preference ranking P_i if $\left|\{o' \in \mathcal{O} : o' P_i o\}\right| = n - 1$. In that case, we also say that i ranks o in the n-th position, or that o has the n-th position in i's preference ranking. We use the term 'higher rank' and 'better position' to refer to a position corresponding to a smaller n.

We start by describing the (individual-proposing) *Deferred Acceptance (DA)* mechanism of Gale and Shapley (1962). Then, we provide a novel and helpful way to instrumentalize the DA to define a class of mechanisms. For this class, we provide a general characterization result for having problem-specific and worst-case auditability index of two.

Deferred Acceptance (DA) for input θ***:*** *Start with Step 1.*

Step $t \geq 1$*. Each individual* i *claims her most preferred object according to* P_i *among those that have not rejected her. Each object* o *that is claimed by some individuals, is tentatively assigned to the claimant with the highest priority score at* o *and rejects the rest. If there are no rejections, the algorithm terminates and the tentative assignments are finalized. Otherwise, we proceed to Step* $t + 1$*.*

Given θ, denote the outcome of the DA mechanism by $DA(\theta)$.

We define a large class of mechanisms that have a DA-representation through some modified priorities. For a given problem $\theta = (P, r)$, we first modify the priority scores by a mapping $\tau : \Theta \to (\mathbb{R}^N)^N$, and then we compute the outcome of the mechanism by applying DA to the modified problem $(P, \tau(\theta))$.

We restrict attention to a mapping $\tau : \Theta \to (\mathbb{R}^N)^N$ that satisfy certain properties. Consider arbitrary two problems $\theta = (P, r)$ and $\theta' = (P', r')$, and denote $\hat{r} := \tau(\theta)$ and $\hat{r}' := \tau(\theta')$. Let $i, j \in \mathcal{I}$ and $o \in \mathcal{O}$ be arbitrary. Then, the following conditions hold.

- **Independence of Irrelevant Alternatives.** Suppose (i) o has the same position in the preference rankings P_i and P'_i, (ii) o has the same position in the preference rankings P_j and P'_j, and (iii) $r_{io} > r_{jo} \iff r'_{io} > r'_{jo}$. Then,

$$\hat{r}_{io} > \hat{r}_{jo} \iff \hat{r}'_{io} > \hat{r}'_{jo}.$$

- **Monotonicity.** Suppose (i) o has a weakly better position in the preference ranking P'_i compared to P_i, (ii) o has a weakly worse position in the preference ranking P'_j compared to P_j, and (iii) $r_{io} > r_{jo} \implies r'_{io} > r'_{jo}$. Then,

$$\hat{r}_{io} > \hat{r}_{jo} \implies \hat{r}'_{io} > \hat{r}'_{jo}.$$

- **Equal Treatment.** Suppose (i) o has the same position in preference rankings P_i and P_j, and (ii) $r_{io} > r_{jo}$. Then,

$$\hat{r}_{io} > \hat{r}_{jo}.$$

Definition 3. *We say a mechanism φ is* ***DA-representable****, if there is a mapping τ satisfying the three assumptions above, such that for any problem $\theta = (P, r)$,*

$$\varphi(\theta) = DA(P, \tau(\theta)).$$

It is straightforward that DA itself is a DA-representable mechanism: for each $i \in \mathcal{I}$, we can take τ to be the 'identity' projection of priority scores. That is, for any $\theta = (P, r)$, let $\tau(\theta) = r$.

More interestingly, the widely applied **Immediate Acceptance (IA)** mechanism is also DA-representable.

Immediate Acceptance (IA) for input θ: *Start with Step 1.*

Step *$t \geq 1$. Each available individual claims her most preferred available object. Each object that is claimed by some individuals, is assigned to the the claimant with the highest priority score for the object. The assigned objects become unavailable. If there are no available objects, the algorithm terminates. Otherwise, we proceed to Step $t + 1$.*

Given θ, denote the outcome of the IA mechanism by $IA(\theta)$.

Consider the following mapping τ. For any problem $\theta = (P, r)$, the modified priority scores $\hat{r} = \tau(\theta)$, and all $i, j \in \mathcal{I}, o \in \mathcal{O}$,

$$\hat{r}_{io} > \hat{r}_{jo}$$

if and only if either o is ranked in a strictly better position at P_i than at P_j, or o is ranked in the same position at P_i and P_j, and $r_{io} > r_{jo}$. Then, IA is DA-representable through the mapping above.

More generally, the family of **Application-Rejection (AR)** mechanisms of Chen and Kesten (2017) are examples of DA-representable mechanisms and include both IA and DA as special cases. For a given natural number $e \in \mathbb{N}$, let AR_e denote the AR mechanism corresponding to parameter e. The standard description of these mechanisms are provided in Appendix B. In the following, we describe the mechanisms through a DA-representation.

Fix an $e \in \mathbb{N}$. We say that o is a *tier* $t \in \mathbb{N}$ object for i if $e(t-1) \leq \left|\left\{o' \in \mathcal{O} : o' P_i o\right\}\right| < et$. Consider the following mapping τ_e. For any problem $\theta = (P, r)$, the modified priority scores $\hat{r} = \tau(\theta)$, and all $i, j \in \mathcal{I}, o \in \mathcal{O}$,

$$\hat{r}_{io} > \hat{r}_{jo}$$

if and only if i ranks o in a better (more preferred) tier than j, or i and j rank o in the same tier and $r_{io} > r_{jo}$. We define AR_e as the mechanism that is DA-representable through τ_e.

Note that when $e = 1$, AR_e is equivalent to IA, and for any $e \geq N$, AR_e is equivalent to DA. When $2 \leq e < N$, AR_e is known as a *Chinese parallel* mechanism (Chen and Kesten, 2017).

The class of DA-representable mechanisms includes a large subclass (but not all) of the Preference Rank Partitioned rules studied by Ayoade and Pápai (2023). In particular, DA-representable rules cover all the ARs, the Secure Immediate Acceptance mechanism (Dur, Hammond, and Morrill, 2019), the class of First-Priority-First mechanisms (Pathak and Sönmez, 2013), and French-priority mechanisms Bonkoungou (2017).

Before stating our main general result in this section, we introduce one more definition. For a given (P, r), we say that an outcome ω is **stable** if there are no two individuals $i, j \in \mathcal{I}$ and object $o \in \mathcal{O}$ such that (1) $oP_i\omega(i)$, (2) $\omega(j) = o$, and (3) $r_{io} > r_{jo}$. Here is an important implication of the DA-representation: if a φ is DA-representable through a mapping τ, then for any problem $\theta = (P, r)$, then the outcome $\varphi(P)$ is stable at problem $(P, \tau(\theta))$. This observation follows directly from the definition of DA-representability and stability of the DA (Gale and Shapley, 1962).

In the following theorem, we fully characterize problems where a given DA-representable mechanism has an auditability index of two, as well as all DA-representable mechanisms with a worst-case auditability index of two.

Theorem 2. *Consider an arbitrary mechanism φ, which is DA-representable through the mapping τ.*

1. *For a given problem $\theta = (P, r)$, $\#\varphi^\theta = 2$ if and only if for any outcome $\omega \neq \varphi(\theta)$, either ω is not stable at $(P, \tau(\theta))$, or it is stable at $(P, \tau(\theta))$, and there are two individuals i and j, such that*
    - *i and j prefer each other's objects at ω more than their own ones, that is, $\omega(j)P_i\omega(i)$ and $\omega(i)P_j\omega(j)$,*
    - *for any $o \notin \{\omega(i), \omega(j)\}$, $oP_i\omega(i)$ or $oP_j\omega(j)$.*
2. *$\#\varphi = 2$ if and only if the conditions in point 1 are satisfied for all problems.*

Proof. See Appendix A.3. □

To put it simply, Theorem 2 says that a DA-representable mechanism will have an auditability index of two only when either stable outcomes are unique or all other stable outcomes are sufficiently undesirable for individuals.

For any DA-representable mechanism φ and any problem $\theta = (P, r)$, the characterization result gives a simple recipe for verifying whether $\#\varphi^\theta = 2$ or not, in polynomial time. Instead of considering all stable outcomes at $(P, \tau(\theta))$, let us just compute the stable outcome $\bar{\omega}$ that is least preferred by all individuals (this can be computed by the object-proposing version of the DA (Gale and Shapley, 1962)). Note that $\bar{\omega}$ would satisfy the two preference conditions in Theorem 2 if and only if all stable outcomes $\omega \neq \varphi(\theta)$ at $(P, \tau(\theta))$ would satisfy them. Hence, to verify that $\#\varphi^\theta = 2$, one can just check whether $\bar{\omega}$ satisfies the two preference conditions or not. Of course, this can be done computationally efficiently.

Theorem 2 has other important corollaries. First, we can use the result to compute the worst-case auditability index of IA.

Proposition 2. $\#IA = 2$.

Proof. By Theorem 2, it is sufficient to show that for any problem $\theta = (P, r)$ there is a unique stable outcome at problem $(P, \tau(\theta))$, where τ is the mapping in the DA-representation of IA.

Consider an arbitrary problem $\theta = (P, r)$, and let $\hat{r} = \tau(\theta)$. To show that there is a unique stable outcome at problem $(P, \hat{r})$, we prove the following equivalent claim.

Claim. At every step t of the implementation of IA, if an individual i receives an object o, then i receives o at every stable outcome of problem $(P, \hat{r})$.

We prove the claim by induction. Suppose i receives o at step 1 of the implementation of IA. Then, o is the most preferred object of i, and she has the highest original priority score at o among all individuals for whom o is the most preferred object. Hence, by construction of $\hat{r}$, individual i has the highest modified priority score at o, i.e., $i = \arg\max_{j \in \mathcal{I}} \hat{r}_{jo}$. Thus, i must receive o at every stable outcome of problem $(P, \hat{r})$.

Now suppose the claim holds for steps $1, 2, \ldots, t-1$, and i receives o at step t of the implementation of IA. Then, o is the t-th most preferred object of i according to P_i, and there is no other available individual $j \neq i$ who ranks o in a higher position than t on P_j. Hence, by construction of $\hat{r}$, individual i has the highest modified priority at o among all available individuals. Thus, i receives o at every stable outcome of problem $(P, \hat{r})$. This completes the proof of the claim, and therefore the proof of Proposition 2. $\square$

The next direct corollary of Theorem 2 is a full characterization of all problems where DA has an auditability index of two.

Proposition 3. *$\#DA^\theta = 2$ if and only if for any $\omega \neq DA(\theta)$, either ω is not stable at θ, or it is stable at θ and there are two individuals i and j, such that*

- *i and j prefer each other's objects at ω more than their own ones, that is, $\omega(j)P_i\omega(i)$ and $\omega(i)P_j\omega(j)$,*
- *for any $o \notin \{\omega(i), \omega(j)\}$, $oP_i\omega(i)$ or $oP_j\omega(j)$.*

This gives a relatively simple characterization of all problems for which DA has an auditability index of two. The necessary conditions for having $\#DA^\theta = 2$ are generally restrictive. One can use the characterization result to compute the proportion of problems for which DA achieves an auditability index of two, either computationally or analytically. In our next result, we show that the worst-case auditability index of DA is maximally large. We prove this as a part of a more general proposition.

Proposition 4. *$\#AR_e = N$ for any $e > 1$. In particular, $\#DA = N$.*

Proof. Consider an arbitrary AR_e with $e > 1$. It is sufficient to find a problem θ, such that $\#AR_e^\theta = N$.

Suppose individuals and objects are indexed, i.e., $\mathcal{I} = \{i_1, \ldots, i_N\}$ and $\mathcal{O} = \{o_1, \ldots, o_N\}$, and consider a problem θ where for all $n \in \{1, \ldots, N\}$, i_n ranks o_{n+1} as her most preferred object, and she ranks o_n as her second most preferred object. Here, we define $o_{N+1} := o_1$. These preferences are illustrated in the table below.

i_1	i_2	$\cdots$	i_{N-1}	i_N
o_2	o_3	$\ldots$	o_N	o_1
o_1	o_2	$\ldots$	o_{N-1}	o_N
$\ldots$	$\ldots$	$\ldots$	$\ldots$	$\ldots$

Moreover, suppose that for all $n \in \{1, \ldots, N\}$, i_n has the highest priority at o_n.

In that problem, AR_e assigns every individual her most preferred object, i.e.,

$$AR_e(\theta) = \left\{(i_n, o_{n+1})\right\}_{n=1}^{N}.$$

Consider the outcome $\omega = \left\{(i_n, o_n)\right\}_{n=1}^{N}$ and an arbitrary subset of individuals $I \subset \mathcal{I}$ (i.e. $|I| < N$). It suffices to show that I does not detect the deviation $\omega \neq AR(\theta)$. Take an arbitrary $i_n \notin I$, and let θ'_{i_n} differ from θ_{i_n} by letting i_n rank o_n as her first choice. Then, $AR_e(\theta_{-i_n}, \theta'_{i_n}) = \omega$. This completes the proof. □

Together, Propositions 2 and 4 reveal a strong contrast among the AR mechanisms: namely, $\#AR_1 = 2$, and $\#AR_e = N$ for all $e > 1$.

4.3 House Allocation

In this section, we study the house allocation model (Hylland and Zeckhauser, 1979) which carries both theoretical importance and covers a wide range of applications such as the assignment of public houses, dormitory rooms, transplant organs and more. For each $i \in \mathcal{I}$, a type report θ_i now takes the form of a strict preference ranking over objects $\mathcal{O}$.

We concentrate on the large class of *sequential dictatorships* that have first been studied by Ehlers and Klaus (2003), Pápai (2001) and Papai (2000). Let a suboutcome σ be a one-to-one function from $I \subset \mathcal{I}$ to $O \subset \mathcal{O}$ with $|I| = |O|$. Denote $\sigma^{\emptyset}$ as the empty suboutcome and let S' be the set of all suboutcomes. Given a suboutcome $\sigma \in S'$, let $\bar{I}(\sigma) := \mathcal{I} \setminus I$ and $\bar{O}(\sigma) := \mathcal{O} \setminus O$ be the set of individuals and objects that are unassigned under σ, respectively. A **dictatorial structure** is a mapping $\pi : S' \to \mathcal{I}$ such that for any suboutcome σ, $\pi(\sigma) \in \bar{I}(\sigma)$. We also call $\pi(\sigma)$ the dictator at σ.

Definition 4. *A mechanism φ is a **sequential dictatorship** if there is a dictatorial structure π such that for each θ, the outcome of φ can be determined through the following algorithm:*

Step 1. *Let* $\sigma^{\emptyset} = \sigma_0(\theta)$. *Individual* $i_1(\theta) \equiv \pi(\sigma^{\emptyset}) = i_d$ *is the first dictator. She chooses her favorite object* $o_1(\theta)$ *from the set of unassigned objects. Let* $\sigma_1(\theta)$ *be the suboutcome/outcome in which* $i_1(\theta)$ *is assigned to* $o_1(\theta)$ *and all other individuals are unassigned. If all individuals have been assigned an object, stop. If not, continue to Step* 2.

Step $t \geq 2$. *The individual* $i_t(\theta) \equiv \pi(\sigma_{t-1}(\theta))$ *is the* t*-th dictator. She chooses her favorite object* $o_t(\theta)$ *from all unassigned objects. Let* $\sigma_t(\theta)$ *be the suboutcome/outcome that assigns individuals according to* $\sigma_{t-1}(\theta)$ *and additionally assigns* $i_t(\theta)$ *to* $o_t(\theta)$*. If all individuals have been assigned an object, stop. If not, continue to Step* $t+1$.

In words, at each step, the respective dictator is assigned to her most preferred object still left after all previous dictators have been assigned. In this sense, for each problem, the algorithm sequentially assigns individuals according to a strict *dictatorial ordering* over individuals. Moreover, the first dictator is always the same and the identity of dictator at a step only depends on the assignments of previous dictators and not on their detailed type reports. Also note that for each problem the algorithm terminates after exactly N steps.

Prior to presenting our first result, we need to introduce some additional notation: Given any sequential dictatorship φ and for each step $t \leq N$, let $I_t^{\varphi} \subseteq \mathcal{I}$ be the set of individuals that are dictators at step t for at least one problem. Let S^{φ} contain all suboutcomes that can realize at some step of the inducing algorithm. A prominent special case of a sequential dictatorship is the well-known *serial dictatorship* (Satterthwaite and Sonnenschein, 1981) defined as follows. A sequential dictatorship φ is a **serial dictatorship**, if $|I_t^{\varphi}| = 1$ for all $t \in \{1, \ldots, N\}$. That is, individuals choose according to a fixed dictatorial ordering.

Our first result establishes that sequential dictatorships are not easy to audit in general.

Proposition 5. *Let* $N \geq 4$. *Then, there exists a sequential dictatorship* φ *with* $\#\varphi \geq N-1$.

Proof. Consider the sequential dictatorship φ specified as follows: For all $t \in \{1, \ldots, N-2\}$, let $|I_t^\varphi| = 1$. Denote $\{i, j\} := I_{N-1}^\varphi \cup I_N^\varphi$ and let $I_{N-1}^\varphi \cap I_N^\varphi \neq \emptyset$. Thus, there exists a problem θ' such that $\pi(\sigma_{N-2}(\theta')) = i$. Let $\Theta' := \{\theta'' \in \Theta | \ \sigma_{N-2}(\theta'') = \sigma_{N-2}(\theta')\}$ and assume that for all $\theta^* \notin \Theta'$, we have $\pi(\sigma_{N-2}(\theta^*)) = j$. In the following, let $o', o'' \in \bar{O}(\sigma_{N-2}(\theta'))$.

We proceed with constructing the problem which will be the basis for the deviation. Consider a problem θ such that

- $o' \ \theta_i \ o''$ and $o' \ \theta_j \ o''$ and for all $\hat{o} \notin \{o', o''\}$, we have $\hat{o} \ \theta_i \ o'$ and $\hat{o} \ \theta_j \ o'$,
- $\varphi(\theta')(i')$ ranks first on $\theta_{i'}$ for all $i' \in \mathcal{I} \setminus \{i, j\}$.

A quick inspection reveals that for all $i' \in \mathcal{I} \setminus \{i, j\}$, it must be $\varphi(\theta')(i') = \varphi(\theta)(i')$. Hence, $\theta \in \Theta'$, which implies that $\varphi(\theta)(i) = o'$ and $\varphi(\theta)(j) = o''$.

Let $\omega \neq \varphi(\theta)$ be a deviation at problem θ, where

- $\omega(i) = \varphi(\theta)(j)$ and $\omega(j) = \varphi(\theta)(i)$,
- $\omega(i') = \varphi(\theta)(i')$, for all $i' \in \mathcal{I} \setminus \{i, j\}$.

Consider an arbitrary $I \subseteq \mathcal{I}$ with $|I| = N - 2$. It suffices to show that I does not detect the deviation. Since $N \geq 4$, we must be in one of the following four cases:

In the first case, suppose that $I \cap \{i, j\} = \emptyset$. It is clear that since $\omega(i') = \varphi(\theta')(i')$ for all $i' \in \mathcal{I} \setminus \{i, j\}$, individuals I cannot detect ω. In the second case, suppose that $j \notin I$ and $i \in I$. Thus, there exists $k \notin I$ with $k \neq j$. Consider a problem $\hat{\theta} = (\hat{\theta}_{\{j,k\}}, \theta_{-\{j,k\}})$ where k ranks o' first on $\hat{\theta}_k$ and j ranks $\varphi(\theta)(k)$ first on $\hat{\theta}_j$. This leads to $\omega(k') = \varphi(\hat{\theta})(k')$ for all $k' \in I$ and hence I does not detect ω. In the third case, suppose that $i \notin I$ and $j \in I$. Consider a problem $\tilde{\theta} = (\tilde{\theta}_i, \theta_{-i})$, where i ranks o'' first on $\tilde{\theta}_i$. Then, it must be $\varphi(\tilde{\theta}) = \omega$, which means that I does not detect ω. In the fourth case, let $\{i, j\} \subset I$. Since $N \geq 4$, we can select two individuals $l, l' \notin I$. Consider a problem $\bar{\theta} = (\bar{\theta}_{\{l,l'\}}, \theta_{-\{l,l'\}})$ such that l ranks $\varphi(\theta)(l')$ first on $\bar{\theta}_l$ and l'

ranks $\varphi(\theta)(l)$ first on $\bar{\theta}_{l'}$. Then, $\varphi(\bar{\theta})(k') = \omega(k')$ for all $k' \in I$. Thus, I does not detect ω. We thus conclude that $\#\varphi \geq N-1$, when $N \geq 4$. $\square$

The constructed mechanism in the proof works like a *serial dictatorship* including step $N-2$ and the individuals i and j always choose at the last two steps. Furthermore, in the constructed mechanism only for one particular suboutcome σ, where i and j are both still unassigned, individual i is the dictator. We then focus on a problem where we reach this particular suboutcome σ and where i and j compete for one of the remaining two objects. The deviation ω at this problem only swaps the assignments of i and j. Intuitively, any group that wishes to detect this deviation must be able to verify that σ was reached and that i is not receiving her favorite remaining object. However, such a verification would require a group of at least $N-1$ individuals.

Yet, some sequential dictatorships can still achieve near maximal auditability. Before stating the main characterization that formalizes this insight, we introduce a new class of mechanisms that will be central for the result.

Definition 5. *A sequential dictatorship φ is a **vice dictatorship** if the following three conditions are satisfied:*

(1) I_t^φ is a single individual for each $t \in \{4, \ldots, N-2\}$.

(2) There are at most two individuals in $I_2^\varphi \cup I_3^\varphi$.

(3) For any pair $\sigma, \hat{\sigma} \in S^\varphi$, if $\bar{O}(\sigma) = \bar{O}(\hat{\sigma})$ and $\bar{I}(\sigma) = \bar{I}(\hat{\sigma})$, then $\pi(\sigma) = \pi(\hat{\sigma})$.

In other words, there are at most two individuals that share the roles as the second and third dictator, while for all other steps except the last two steps, only one individual can be the dictator.[12] Additionally, condition (3) requires that at each step, the identity of the dictator is fully determined by the set of still unassigned objects and unassigned individuals.[13]

[12]In fact, condition (1) and (2) together imply that there are at most two individuals that can be dictators at the last two steps, i.e. $|I_{N-1}^\varphi \cup I_N^\varphi| = 2$.

[13]Clearly, condition (1) and (2) already imply that condition (3) holds for any pair of suboutcomes until step $N-1$.

Also note that the class of vice dictatorships contains all serial dictatorships. Our main result of this section establishes that vice dictatorships characterize all sequential dictatorships with a worst-case auditability index of two.

Theorem 3. *Let $|N| \geq 5$. If φ is a sequential dictatorship, then $\#\varphi = 2$ if and only if φ is a vice dictatorship.*

Proof. See Appendix A.4 □

Essentially, when φ is not a vice dictatorship, then one can identify two individuals i, j whose relative position in the dictatorial orderings changes across two problems $\hat{\theta}$, $\tilde{\theta}$ while the same two objects o', o'' remain unassigned until either i or j is at turn. Moreover, at these two problems all other individuals rank o', o'' at the last two positions and receive their best ranked objects, whereas i and j both rank o' and o'' at the first and second position, respectively. The key features of the two problems basically ensure that no pair of individuals can unveil whom of i and j should have received o'. Similar as in the proof of Proposition 5, we therefore can construct a deviation with respect to one of these two problems that only swaps the assignments of i and j and cannot be detected by any pair of individuals.

By contrast, vice dictatorships can be regarded as 'almost' serial dictatorial. Specifically, if there is an indeterminacy of relative positions in the dictatorial orderings for two individuals that these two individuals cannot resolve with their information as a group, then one of these individuals can resolve the indeterminacy together with the first dictator. This eventually implies that there are always two individuals that can detect a deviation. More generally, in conjunction with Proposition 5 the result indicates that pursuing a high level of auditability requires dictatorial structures with low complexity.

5 Other Applications

5.1 Single-Object Auctions

Consider the problem of selling a single object to individuals in $\mathcal{I}$. The setup corresponds to the special case of our problem with $\Omega_i = \{0,1\} \times \mathbb{R}_+$ for all $i \in \mathcal{I}$. Given an outcome $\omega \in \times_{i\in\mathcal{I}}\Omega_i$, we interpret $\omega(i) = (1, y)$ as i receiving the object and paying y, and $\omega(i) = (0, y)$ as i not receiving the object and paying y. The space of feasible outcomes $\Omega \subseteq \times_{i\in\mathcal{I}}\Omega_i$ is such that exactly one individual receives the object. A problem $\theta = (b_i)_{i\in\mathcal{I}} \in \mathbb{R}_+^{\mathcal{I}}$ is a vector of individuals' bids such that $b_i \neq b_j$ for any $i, j \in \mathcal{I}$.[14] Let Θ be the space of all problems.

We say an auction mechanism is a **fixed-pay auction**, if each individual's payment only depends on her bid and on whether she receives the object or not. Formally, a fixed-pay auction is a mechanism $\varphi : \Theta \to \Omega$ for which there are no problems $b, b' \in \Theta$ and an individual $i \in \mathcal{I}$ with $b_i = b'_i$, such that for $(x, y) := \varphi(b)(i)$ and $(x', y') := \varphi(b')(i)$, $x = x'$ and $y \neq y'$. The class of fixed-pay auctions includes two important mechanisms: the first-price auction and the all-pay auction.

The **First-Price Auction (FPA)** is a fixed-pay auction mechanism such that for any problem $b \in \Theta$ and the corresponding outcome $\omega = FPA(b)$,

$$\omega(i) = \begin{cases} (1, b_i) & b_i = \max_{j\in\mathcal{I}} b_j, \\ (0, 0) & \text{otherwise.} \end{cases}$$

The **All-Pay Auction (APA)** is a fixed-pay auction mechanism such that for any problem $b \in \Theta$ and the corresponding outcome $\omega = APA(b)$,

$$\omega(i) = \begin{cases} (1, b_i) & b_i = \max_{j\in\mathcal{I}} b_j, \\ (0, b_i) & \text{otherwise.} \end{cases}$$

[14] Unequal bids are necessary for defining the common auction mechanisms. We may think that it is unlikely to have problems where some individuals submit the exact same bids. Alternatively, we may assume that there is a deterministic (observable) tie-breaking rule.

Our first result is a characterization of the entire class of auctions mechanisms with a worst-case auditability index of one. We say a fixed-pay auction φ is a **dual-dictatorship**, if there are individuals i_1 and i_2, and a subsets of bids $\bar{B} \subseteq \mathbb{R}_+$, such that for any problem $b \in \Theta$, and the corresponding outcomes $(x_1, y_1) = \varphi(b)(i_1)$ and $(x_2, y_2) = \varphi(b)(i_2)$, the following conditions hold:

(1). $x_1 = 1$ if and only if $b \in \bar{B}$,

(2). $x_2 = 1$ if and only if $b \notin \bar{B}$.

In other words, dual-dictatorships require that only two individuals i_1 and i_2 can receive the object, and moreover, i_1 receives the object whenever her bid is in some set $\bar{B}$, and otherwise, i_2 receives that object.

Theorem 4. *$\#\varphi = 1$ if and only if the auction mechanism φ is a dual-dictatorship.*

Proof. See Appendix A.5. ◻

From Theorem 4, we conclude that the auditability indices of the FPA and APA shall be weakly larger than two. In our next result, we show their auditability indices are exactly two, which means that the mechanisms are maximally auditable among non-dual-dictatorial fixed-pay auctions.

Proposition 6. *$\#FPA = \#APA = 2$.*

The result is immediate given the following observation: under both auction formats, whenever an individual i with the highest bid does not receive the object, and some individual j with a lower bid receives it, then i and j detect the deviation. The formal proof is immediate and therefore omitted.

The final (non-fixed-pay) auction mechanism that we study is the **Second-Price Auction (SPA)**, defined as follows: for any problem $b \in \Theta$ and the corresponding outcome $\omega = SPA(b)$,

$$\omega(i) = \begin{cases} (1, \bar{b}) & b_i = \max_{j \in \mathcal{I}} b_j, \text{ and} \\ (0, 0) & \text{otherwise,} \end{cases}$$

where $\bar{b} = \max_{j\in\mathcal{I}\setminus\{i\}} b_j$ is the second highest bid. We show that the SPA is maximally not auditability for any problem.

Proposition 7. *For any problem $b \in \Theta$, $\#SPA^b = N$.*

Consider a deviation where an individual i with the highest bid receives the object and pays a bid $\tilde{b}$ that is strictly larger than the second highest bid $\bar{b}$, but smaller than b_i. Then, no proper subset of individuals can detect the deviation, as they cannot know whether someone outside of the subset has bid $\tilde{b}$ or not. The formal proof is immediate and therefore omitted.

5.2 Voting with a Binary Outcome

Consider the problem where individuals $\mathcal{I}$ collectively choose to implement one of two alternatives. This corresponds to the special case of our problem where $\Omega_i = \{0,1\}$ for all $i \in \mathcal{I}$, and the set of feasible outcomes $\Omega \subseteq \{0,1\}^{\mathcal{I}}$ is such that $\omega(i) = \omega(j)$ for all $\omega \in \Omega$ and $i,j \in \mathcal{I}$. Let $\theta_i \in \{0,1\}$ for all $i \in \mathcal{I}$.

We say a social choice mechanism is **dictatorial** if there is an individual $i \in \mathcal{I}$, such that $\varphi(\theta) = \varphi(\theta')$ for any problems θ and θ' with $\theta_i = \theta'_i$.

Theorem 5. $\#\varphi = 1$ *if and only if the social choice mechanism φ is dictatorial.*

Proof. See Appendix A.6. ☐

The dictatorial mechanism gives all the decision power to a single individual. It may be interesting to study mechanisms that treat individuals 'symmetrically', which we do next. For the rest of this section, we assume that $N = |\mathcal{I}|$ is an odd number. We say a social choice mechanism φ is **anonymous**, if $\varphi(\theta) = \varphi(\theta')$ for any $\theta \in \Theta = \{0,1\}^{\mathcal{I}}$, and θ' can be obtained from θ by permuting zeros and ones. In other words, an anonymous mechanism only considers the numbers of zeros and ones in the problem, and not the individuals' identities.

An important example of an anonymous social choice mechanism is the *majority voting* mechanism. A social choice mechanism is a **majority voting mechanism** for outcome $x \in \{0,1\}$, if for any $\theta \in \Theta = \{0,1\}^{\mathcal{I}}$,

$$\varphi(\theta)(i) = (x, x, \ldots, x) \text{ if and only if } \sum_{i \in \mathcal{I}} \theta_i \geq \frac{N+1}{2}.$$

It is easy to see that the auditability index of a majority voting mechanism is $(N+1)/2$. In the following theorem, we give a stronger result: we characterizes majority voting as the unique most auditable anonymous social choice mechanism.

Theorem 6. *If φ is an anonymous social choice mechanism, then $\#\varphi \geq \frac{N+1}{2}$. Moreover, $\#\varphi = \frac{N+1}{2}$ if and only if φ is the majority voting mechanism.*

Proof. The proof that a majority voting mechanism has a worst-case auditability index of $(N+1)/2$ is immediate, and therefore omitted. To prove the theorem, it is sufficient to show that any other anonymous mechanism has a worst-case auditability index of strictly larger than $(N+1)/2$.

Consider an anonymous social choice mechanism φ, which is not the majority voting mechanism. Then, there are problems θ and θ' with either

$$n := \sum_{i \in \mathcal{I}} \theta_i < \sum_{i \in \mathcal{I}} \theta'_i := n' \leq \frac{N-1}{2},$$

or

$$n := \sum_{i \in \mathcal{I}} (1 - \theta_i) < \sum_{i \in \mathcal{I}} (1 - \theta'_i) := n' \leq \frac{N-1}{2},$$

such that $\varphi(\theta) = (x, x, \ldots, x) \neq (y, y, \ldots, y) = \varphi(\theta')$. Without loss of generality, (by renaming zeros and ones if necessary) suppose that $n := \sum_{i \in \mathcal{I}} \theta_i < \sum_{i \in \mathcal{I}} \theta'_i := n' \leq \frac{N-1}{2}$. Moreover, again without loss of generality (by choosing a larger n and a smaller n'), suppose that n and n' are consecutive numbers, i.e., $n' = n + 1$.

Consider the deviation $(y, y, \ldots, y) \neq \varphi(\theta)$. We argue that no subset of individuals $I \subseteq \mathcal{I}$ with $|I| < N - n$ can detect this deviation. Since $I \subseteq \mathcal{I}$, it should be that

$$\sum_{i \in I} \theta_i \leq \sum_{i \in \mathcal{I}} \theta_i = n.$$

Therefore, there are type reports $\tilde{\theta}_{-I}$ of individuals in $\mathcal{I} \setminus I$, such that $\sum_{i \in I} \theta_i + \sum_{i \in \mathcal{I} \setminus I} \tilde{\theta}_i = n'$. Namely, this condition is satisfied for $\tilde{\theta}_{-I}$ that differs from θ_I by only that one additional individual in $\mathcal{I} \setminus I$ has a type report equal to one, instead of zero. Since, $\sum_{i \in I} \theta_i + \sum_{i \in \mathcal{I} \setminus I} \tilde{\theta}_i = n' = \sum_{i \in \mathcal{I}} \theta'_i$, and φ is anonymous, we get that

$$\varphi(\theta_I, \tilde{\theta}_{-I}) = \varphi(\theta') = (y, y, \ldots, y).$$

Thus, I does not detect the deviation $(y, y, \ldots, y)$.

Since $I, |I| < N - n$ was arbitrary, by definition of the auditability index we get that

$$\#\varphi^{\theta} \geq N - n > N - \frac{N-1}{2} = \frac{N+1}{2}.$$

This completes the proof of Theorem 6. $\square$

Note that the characterization results in Theorems 5 and 6 crucially rely on using the notion of worst-case auditability. Specifically, some anonymous (and hence non-dictatorial) social choice mechanisms may have an auditability index of one for some (or most) problems. Consider for instance the 'veto mechanism' that implements the outcome 1 unless some individual's type report is 0. Here, one can easily see that for any problem for which at least one individual has a type report equal to 0, the veto mechanism has an auditability index of one. However, also note that for the problem where each individual's type report is 1, the veto mechanism has an auditability index of N.

5.3 Choice with Affirmative Action

We study the problem of selecting a subset of individuals for up to $Q < |\mathcal{I}|$ positions, subject to meeting certain distributional targets. The individuals are partitioned into two (disjoint) subsets I_L and I_H, where I_L denotes the individuals from low-income status, and I_H denotes the individuals of high-income status. The type report $\theta_i = \rho_i \in \mathbb{R}_+$ of individual $i \in \mathcal{I}$ is her priority score. The space of options for each i is $\Omega_i = \{0, 1\}$, where 1 denotes being selected, and 0 denotes not being selected. The set of feasible outcomes $\Omega \subseteq \{0, 1\}^{\mathcal{I}}$ are those ω for which $\sum_{i \in \mathcal{I}} \omega(i) = Q$ and

$\sum_{i \in I_L} \omega(i) \geq R$. The former condition can be interpreted as non-wastefulness of the outcome, and the latter condition can be interpreted as meeting the distributional targets.[15] The model covers a myriad of important applications, including all those mentioned in the introduction.

In these problems, an outcome is typically chosen by **reserves** mechanisms, where low-income applicants are preferentially treated for R of the (reserved) positions. There are two well-studied reserves mechanisms, which we discuss below.

- **Reserved-seats-first (RSF) mechanism:** For a given $\theta \in \Theta$, $REG(\theta)$ is determined as follows:

  In the first step, choose up to R highest priority individuals from I_L, and let $\bar{I}$ denote the set of chosen individuals. In the second step, choose up to $Q - |\bar{I}|$ highest priority individuals from $\mathcal{I} \setminus \bar{I}$. Let I denote the set of chosen individuals after these two steps. Then, for all $i \in \mathcal{I}$,

  $$REG(\theta)(i) = \mathbb{1}[i \in I].$$

- **Open-seats-first (OSF) mechanism:** For a given $\theta \in \Theta$, $OSF(\theta)$ is determined as follows:

  In the first step, choose up to $Q - R$ highest priority individuals from $\mathcal{I}$, and let $\bar{I}$ denote the set of chosen individuals. In the second step, choose up to R highest priority individuals from $I_L \setminus \bar{I}$, and let $\tilde{I}$ denote the set of chosen individuals in this step. In the third step, choose up to $R - |\tilde{I}|$ highest priority individuals from $\mathcal{I} \setminus (\bar{I} \cup \tilde{I})$. Let I denote the set of chosen individuals after these three steps. Then, for all $i \in \mathcal{I}$,

  $$OSF(\theta)(i) = \mathbb{1}[i \in I].$$

[15] Restricting the set of feasible outcomes to those that are non-wasteful and meet the distributional targets is crucial for our results. More specifically, our results do not extend to the environment where the designer could choose a deviation violating these conditions. Our modelling choice is natural in an environment where non-wastefulness and meeting distributional objectives are hard requirements.

There are well-known results about the properties of the RSF and OSF rule. The RSF mechanism minimizes priority violations (i.e., situations where a lower priority individual is chosen, and a higher priority one is not) subject to meeting the reserves (Abdulkadiroğlu and Grigoryan, 2021; Echenique and Yenmez, 2015). On the other hand, the OSF mechanism provides an additional boost to low-income applicants: the mechanism first chooses some of the highest priority low-income applicants from the general pool (step one), and only after that, it chooses an additional R of the remaining (lower priority) low-income applicants (step two). In this way, OSF admits more low-income applicants than RSF (Dur et al., 2018). Choosing more low-income applicants under OSF comes at the cost of creating priority violations, and the same distributional outcomes can be achieved by the RSF with a higher reserves ratio. Abdulkadiroğlu and Grigoryan (2021) mention that the RSF mechanism solves this tradeoff in a more *transparent* way; any priority violation can be explained by diversity objectives under RSF, but not under OSF. Our auditability theory provides a very different angle for comparing the two reserves mechanisms.

Proposition 8. *$\#RSF = R + 2$ and $\#OSF \geq \max\{Q - R + 1, R + 2\}$.*

Proof. See Appendix A.7 $\square$

If the proportion of reserves R/Q is small, there may be a large gap between the auditabilities of RSF and OSF. For example, if this proportion is 20%, the RSF will have about four times smaller auditability index than the OSF. Some examples of small reserves ratios in real-life problems include allocation of H1-B visas with about 23% (Pathak et al., 2020), and pandemic rationing with about 10-40% (Pathak et al., 2021).[16]

[16] A more detailed account of the reserves policy pandemic rationing can be found here: https://www.covid19reservesystem.org/policy-impact.

6 Discussion

In this paper, we introduce and analyze a novel informational property/measure for mechanisms. Namely, we quantify the size of information needed to detect deviations from a mechanism. Our auditability measure proves tractable and informative, and our approach appears insightful in different applications. We believe that there is a scope for future research with refined auditability notions that also account for some important strategic considerations. We discuss some if these possible extensions below.

1. First, in our model, we take type reports (problems) as given, whereas in reality they may be determined endogenously. More specifically, the mechanism itself could influence the realized problem, which in turn will impact the auditability of the mechanism.

2. Second, our auditability index is defined with respect to adversarial deviations from the mechanism (i.e., those that are hardest to detect). Yet, in some applications some deviations may be more likely to occur than others. For example, the mechanism might be implemented by a strategic authority with preferences over possible outcomes, who chooses deviations to maximize their utility. In such a setup, a credibility notion in the spirit of Akbarpour and Li (2020) may be more relevant.

3. Finally, we define detecting a deviation as a situation where a group of individuals can rule out that the promised mechanism was used based on the collective information on the group's type reports and their final assignments. It may be interesting to consider the strategic aspects of how this collective information is generated. For example, if this information about the group is being shared voluntarily, we may restrict attention to detections where everyone in the group prefers the promised outcome to the deviation. Similarly, if the group's information is being observed by some other individual (from or outside of the group), potentially against the group's will, then we may restrict attention to detections where at least one individual prefers the promised outcome to the

deviation. One could also imagine that the group's information is audited by a trusted third party. In this case, one may restrict attention to detections that are optimally planned. That is, the auditing entity may anticipate the likely deviations by the designer and then chooses the subset of observations accordingly.

Our main motivation for keeping our framework general and abstract is to cover a comprehensive spectrum of scenarios, applications and considerations. For example, in many applications there is no centralized authority or the authority is not strategic. As motivated in our introduction, deviations may even result from documentation errors, or other kinds of mistakes and misunderstandings by the operator. How information can be collected or accessed can also vary heavily across applications. Therefore, we believe that our high level analysis is a crucial benchmark for studying auditability. We hope that future research can incorporate strategic considerations with refined auditability notions tailored towards applications where they appear particularly relevant.

Although our analysis abstracts away from certain important features (such as how information may be accessed), our theory can be directly applied to obtain insights for alternative and more specific (microfounded) auditability notions. To support this claim, consider a setup where audits are conducted by a third party, which samples and observes $M \leq N$ of the type reports and outcomes uniformly at random. In this framework, we can measure auditability as the probability of detecting deviations. Suppose we want to understand how Immediate Acceptance and Deferred Acceptance would compare with respect to this measure. From our theory, we know that the Immediate Acceptance mechanism has a worst-case auditability index of two (Proposition 2). Hence, for any problem and any deviation, there are at least two individuals whose information alone is sufficient to detect the deviation. In a sufficiently large market, the probability that the auditor will access both individual's information is at least about $\left(\frac{M}{N}\right)^2$. Hence, $\left(\frac{M}{N}\right)^2$ is a lower bound for the probability of detecting an adversarial deviation. If the audit is strategic instead of uniform random, the detection probability will be even larger. In contrast, since the Deferred Acceptance mechanism has a worst-case auditability index of N (Proposition 4), unless $M = N$, at some problems the probability of detecting a deviation is zero for any audit sample of size $M < N$.

7 Conclusion

We introduce a novel framework for studying auditability of mechanisms and provide new insights on prominent mechanisms in different settings such as allocation problems, auctions, voting, and choice with affirmative action. Our theory and findings may contribute to discussions on the choice of a mechanism for real-life problems.

Our auditability notion might be useful in other problems and domains not studied in this paper. For example, one could explore auditability of mechanisms in related allocation problems such as multi-unit auctions, many-to-many matching problems, or non-binary voting settings. Additionally, one could investigate algorithmic implementations and computational complexity questions related to detecting deviations and measuring auditability. We leave this and related questions, as well as many possible extensions of our theory, for future research.

References

Abdulkadiroğlu, A. and A. Grigoryan (2021): "Priority-Based Assignment with Reserves and Quotas," No. w28689. National Bureau of Economic Research.

Abdulkadiroğlu, A. and T. Sönmez (2003): "School Choice: A Mechanism Design Approach," *American Economic Review*, 93, 729–747.

Acquisti, A., C. Taylor, and L. Wagman (2016): "The economics of privacy," *Journal of economic Literature*, 54, 442–492.

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

Arnosti, N. (2020): "Deferred Acceptance is Unpredictable," Blog Post, https://nickarnosti.com/blog/da-predictability/.

AYGUN, O. AND I. BÓ (2021): "College Admission with Multidimensional Privileges: The Brazilian Affirmative Action Case," *American Economic Journal: Microeconomics*, 13, 1–28.

AYOADE, N. AND S. PÁPAI (2023): "School Choice with Preference Rank Classes," *Games and Economic Behavior*, 137, 317–341.

BENNER, M. AND U. BOSER (2018): "Expanding Access to High Quality Schools (CAP Review)," Center for American Progress.

BONKOUNGOU, S. (2017): "University Admissions Reform in France: An Analysis of a New Mechanism," Working Paper.

BRANDT, F. (2001): "Cryptographic Protocols for Secure Second-Price Auctions," *In International Workshop on Cooperative Information Agents: Springer Berlin Heidelberg*, 154–165.

CANETTI, R., A. FIAT, AND Y. A. GONCZAROWSKI (2023): "Zero-Knowledge Mechanisms," arXiv preprint arXiv:2302.05590.

CHEN, Y. AND O. KESTEN (2017): "Chinese College Admissions and School Choice Reforms: A Theoretical Analysis," *Journal of Political Economy*, 125, 99–139.

CHITRA, T., M. V. FERREIRA, AND K. KULKARNI (2023): "Credible, Optimal Auctions via Blockchains," arXiv preprint arXiv:2301.12532.

DUR, U., R. G. HAMMOND, AND T. MORRILL (2019): "The Secure Boston Mechanism: Theory and Experiments," *Experimental Economics*, 22, 918–953.

DUR, U., S. D. KOMINERS, P. A. PATHAK, AND T. SÖNMEZ (2018): "Reserve Design: Unintended Consequences and the Demise of Boston's Walk Zones," *Journal of Political Economy*, 6, 2457–2479.

ECHENIQUE, F. AND B. M. YENMEZ (2015): "How to control controlled school choice," *American Economic Review*, 105, 2679–2694.

EHLERS, L. AND B. KLAUS (2003): "Consistent House Allocation," *Barcelona, Unpublished mimeo.*

FERREIRA, M. V. AND M. S. WEINBERG (2020): "Credible, Truthful, and Two-Round (Optimal) Auctions via Cryptographic Commitments," *In Proceedings of the 21st ACM Conference on Economics and Computation*, 683–712.

GALE, D. AND L. S. SHAPLEY (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–15.

GANGAM, R. R., T. MAI, N. RAJU, AND V. V. VAZIRANI (2023): "A Structural and Algorithmic Study of Stable Matching Lattices of Multiple Instances," arXiv preprint:2304.02590.

GONCZAROWSKI, Y. A. AND C. THOMAS (2023): "Structural Complexities of Matching Mechanisms," arXiv prepint: 2212.08709.

HAFALIR, I. E., M. B. YENMEZ, AND M. A. YILDIRIM (2013): "Effective Affirmative Action in School Choice," *Theoretical Economics*, 8, 325–363.

HAKIMOV, R. AND M. RAGHAVAN (2020): "Transparency in Centralised Allocation: Theory and Experiment," Working Paper.

HAUPT, A. AND Z. HITZIG (2023): "Contextually Private Mechanisms," arXiv preprint arXiv:2112.10812.

HYLLAND, A. AND R. J. ZECKHAUSER (1979): "The Efficient Allocation of Individuals to Positions," *Journal of Political Economy*, 87(2), 293–314.

IMAMURA, K. (2020): "Meritocracy versus diversity," Working Paper.

MÖLLER, M. (2022): "Transparent Matching Mechanisms," Working Paper.

OLLAR, M., M. ROSTEK, AND J. H. YOON (2021): "Privacy in Markets *," Working paper.

PAPAI, S. (2000): "Strategyproof Assignment by Hierarchical Exchange," *Econometrica*, 68, 1403–1433.

PATHAK, P. A. (2017): "What Really Matters in Designing School Choice Mechanisms," in *Advances in Economics and Econometrics, 11th World Congress of the Econometric Society*, ed. by B. Honore, M. Piazessi, A. Pakes, and L. Samuelson, Cambridge University Press.

PATHAK, P. A., A. REES-JONES, AND T. SÖNMEZ (2020): "Immigration Lottery Design: Engineered and Coincidental Consequences of H-1B Reforms," National Bureau of Economic Research, Working Paper No. w26767.

PATHAK, P. A. AND T. SÖNMEZ (2013): "School Admissions Reform in Chicago and England: Comparing Mechanisms by their Vulnerability to Manipulation," *American Economic Review*, 103(1), 80–106.

PATHAK, P. A., T. SÖNMEZ, U. M. ÜNVER, AND B. M. YENMEZ (2021): "Fair Allocation of Vaccines, Ventilators and Antiviral Treatments: Leaving No Ethical Value Behind in Health Care Rationing," In Proceedings of the 22nd ACM

Conference on Economics and Computation.

Pycia, M. and P. Troyan (2023): “A Theory of Simplicity in Games and Mechanism Design,” *Econometrica*, 91, 1495–1526.

Pycia, M. and M. U. Ünver (2022): “Arrovian Efficiency and Auditability in Discrete Mechanism Design,” Working Paper.

Pápai, S. (2001): “Strategyproof and Nonbossy Multiple Assignments,” *Journal of Public Economic Theory*, 3, 257–271.

Satterthwaite, M. A. and H. Sonnenschein (1981): “Strategy-proof allocation mechanisms at differentiable points,” *The Review of Economic Studies*, 48, 587–597.

Schuler, N. (2018): “CPS OIG Uncovers Widespread Admissions Irregularities in K-8 Options for Knowledge Program,” Office of Inspector General, Chicago Board of Education. Press Release, February 21.

Sönmez, T. and B. M. Yenmez (2022): “Affirmative Action in India via Vertical, Horizontal, and Overlapping Reservations,” *Econometrica*, 90, 1143–1176.

Vickrey, W. (1961): “Counterspeculation, Auctions, and Competitive Sealed Tenders,” *The Journal of Finance*, 1, 8–37.

Woodward, K. (2020): “Self-Auditable Auctions,” Working Paper.

A Omitted Proofs

A.1 Proof of Theorem 1

Part 1. Consider an arbitrary $\theta \in \Theta$. Our goal is to show that $\#\varphi^\theta = 1$ if and only if φ has a sequential clinching implementation at θ.

First, we prove the ‘if’ part. Suppose φ is a sequential clinching mechanism. Consider an arbitrary problem θ, and an arbitrary deviation $\omega \neq \varphi(\theta)$. We show that some individual i detects the deviation ω.

Let i be the first individual in the sequential clinching implementation such that $\varphi(\theta)(i) \neq \omega(i) := o$ (that is, who gets the ‘wrong’ object under ω). Let O be the set of available objects at the step that i clinches $\varphi(\theta)(i)$. By the choice of the step, $o \in O$. (since all individuals in previous steps have received the 'correct' objects). By definition of sequential clinching, $\varphi(\theta)(i)$ is the only possible object among O for i at θ_i. Hence, i detects the deviation ω.

We now prove the 'only if' part. Suppose $\#\varphi^\theta = 1$. For an arbitrary θ, we will find a sequence of individuals $(i_1, i_2, \ldots, i_N)$ that would correspond to the sequential clinching implementation of φ at problem θ.

Let $\omega := \varphi(\theta)$. For an individual i, let $O_i(\theta_i)$ denote the set of possible objects for i at θ_i. First, we show that there is an individual i_1 such that $|O_{i_1}(\theta_{i_1})| = 1$. Suppose, for the sake of contradiction, that there is no such individual. Or equivalently, suppose that $|O_i(\theta_i)| \geq 2$ for all $i \in \mathcal{I}$. Consider a directed graph whose vertices are individuals $\mathcal{I}$, and where we have a directed edge from vertex i to vertex j if and only if there is an object $o \in O_i(\theta_i)$ such that $\omega(j) = o$. Since each vertex has an outdegree of at least one, there is a directed cycle. Select one such cycle, and let the individuals trade their objects at ω according to this cycle. The resulting outcome, call it ω', gives a possible object to each individual. Hence, the deviation $\omega' \neq \omega$ cannot be detected by any single individual. This contradicts that $\#\varphi = 1$.

Thus, i_1 is the desired individual in step 1 of the sequential clinching implementation. Let o_1 be $\varphi(\theta)(i_1)$. Consider the restriction of the problem to individuals $\mathcal{I} \setminus \{i_1\}$ and objects $\mathcal{O} \setminus \{o_1\}$. Again, we can use a proof by contradiction to show that there is an individual an i_2 such that $|O_{i_2}(\theta_{i_2}) \setminus \{o_1\}| = 1$.

We construct the rest of the individuals $(i_3, \ldots, i_N)$ in the same way, completing the proof of Part 1.

Part 2. Next, we prove the following: $\#\varphi = 1$ if and only if φ has a sequential clinching implementation at any problem θ, and moreover, the mechanism can be implemented with a sequential clinching order that only depends on the set of available objects at each step (but otherwise does not depend on the problem).

The first part of the previous statement (sentence) is a direct consequence of Part 1. Our goal is to prove the second part of the statement, that is, to find a sequential clinching implementation of φ, where the selection of individuals only depends on the set of available objects at each step. First, we state and prove an intermediate result.

Claim. Let O and I be any set of available objects and individuals that can be achieved by some sequential clinching implementation of φ at some problem θ. Then, there is an individual $i \in I$ such that $\left|O_i(\theta_i') \cap O\right| = 1$ for all θ_i'.

Proof. The proof of the Claim resembles the proof of the 'only if' part of Theorem 1. Suppose, for the sake of contradiction, that O and I can be achieved by some sequential clinching implementation of φ at some problem θ, and for any $i \in I$, there is a type report θ_i' such that $\left|O_i(\theta_i') \cap O\right| \geq 2$. Consider the problem (θ_I', θ_{-I}), and let $\omega := \varphi(\theta_I', \theta_{-I})$. By definition of the sequential clinching implementation, individuals in $\mathcal{I} \setminus I$ are matched to the same objects in $\mathcal{O} \setminus O$ at both problems θ and (θ_I', θ_{-I}). Consider a directed graph whose vertices are individuals I, and where we have a directed edge from vertex i to vertex j if and only if there is an object $o \in O_i(\theta_i) \cap O$ such that $\omega(j) = o$. Since each vertex has an outdegree of at least one, there is a directed cycle. Select one such cycle, and let the individuals trade their objects at ω according to this cycle. The resulting outcome, call it ω', gives a possible object to each individual. Hence, the deviation $\omega' \neq \omega$ cannot be detected by any single individual. This contradicts that $\#\varphi = 1$. $\square$

Now, consider the following sequential clinching implementation of φ for some arbitrary θ. Let the first clincher i_1 be an individual for whom $|O_{i_1}(\theta_{i_1})| = 1$ for all type reports θ_{i_1}'. In general, let the step t clincher i_t be the available individual for whom $\left|O_{i_t}(\theta_{i_t}') \cap O_t\right| = 1$ for all θ_{i_t}', where O_t is the set of available objects at step t. At any step, if there is more than one such available individual, select one with some fixed tie-breaking rule. The sequential clinching order in this implementation only depends on the set of objects. This completes the proof of Theorem 1.

A.2 Proof of Proposition 1

Suppose $\#\varphi = 1$. By the second part of Theorem 1, there is a sequential clinching implementation of φ where the identity of each step's clincher only depends on the set of available objects. Let $(i_1, i_2, \ldots, i_N)$ be the sequence of individuals in this sequential clinching implementation.

Suppose, for the sake of contradiction, that φ has full range.

Claim. For any problem θ, $|O_{i_n}(\theta_{i_n})| \geq n$ for all $n \in \{1, 2, \ldots, N\}$.

Proof. Consider an arbitrary problem θ and an arbitrary $n \in \{1, 2, \ldots, N\}$. We will show that $|O_{i_n}(\theta_{i_n})| \geq n$.

Suppose, for the sake of contradiction, that $|O_{i_n}(\theta_{i_n})| \leq n - 1$. Since φ has full range, there is a problem θ' such that $\varphi(\theta')$ assigns all objects in $O_{i_n}(\theta_{i_n})$ to the first $|O_{i_n}(\theta_{i_n})|$ individuals in $\{i_1, i_2, \ldots, i_{n-1}\}$. Consider the problem $(\theta_{i_n}, \theta'_{-i_n})$. Then, by definition of the sequential clinching implementation, it should be that $\varphi(\theta_{i_n}, \theta'_{-i_n})$ also assigns all objects $O_{i_n}(\theta_{i_n})$ to the first $|O_{i_n}(\theta_{i_n})|$ individuals in $\{i_1, i_2, \ldots, i_{n-1}\}$. Therefore, $\varphi(\theta_{i_n}, \theta_{-i'_n})(i_n) \notin O_{i_n}(\theta_{i_n})$, which is a contradiction. □

We now prove Proposition 1. If $O_{i_2}(\theta_{i_2}) \neq \mathcal{O}$, fix some $o \in \mathcal{O} \setminus O_{i_2}(\theta_{i_2})$. If $O_{i_2}(\theta_{i_2}) = \mathcal{O}$, o can be arbitrary. Since φ has full range, there is a problem $\bar{\theta}$ such that $\varphi(\bar{\theta})(i_1) = o$. Let $\omega := \varphi(\bar{\theta}_{i_1}, \theta_{-i_1})$. Since $|O_{i_2}(\theta_{i_2})| \geq 2$ (by the Claim), and by the choice of o, we have $|O_{i_2}(\theta_{i_2}) \setminus \{o\}| \geq 2$. Also, since for each $n \in \{3, \ldots, N\}$, $|O_{i_n}(\theta_{i_n})| \geq n \geq 3$ (by the Claim), we have that $|O_{i_n}(\theta_{i_n}) \setminus \{o\}| \geq 2$.

Consider a directed graph whose vertices are individuals $\mathcal{I} \setminus \{i_1\}$, and where we have a directed edge from vertex i to vertex j if and only if there is an object $o' \in O_i(\theta_i) \setminus \{o\}$ such that $\omega(j) = o'$. Since each vertex has an outdegree of at least one, there is a directed cycle. Select one such cycle, and let the individuals trade their objects at ω according to this cycle. The resulting outcome, call it ω', gives a possible object to each individual. Hence, the deviation $\omega' \neq \omega$ cannot be detected by any single individual. This contradicts that $\#\varphi = 1$.

A.3 Proof of Theorem 2

The second part of Theorem 2 follows directly from the first part. In what follows we prove the first part of Theorem 2.

Suppose φ is DA-representable through mapping $\tau : \Theta \to (\mathbb{R}^N)^N$. Fix an arbitrary problem $\theta = (P, r)$. To simplify notation, let us denote the modified priorities by $\hat{r} = \tau(\theta)$.

First, we establish that φ does not have a sequential clinching implementation at θ. More specifically, no individual i can clinch an object o at θ. To see this, consider a type report θ'_j of some individual $j \neq i$, where j ranks o as her most preferred object, and she has a strictly higher priority score at o than i. Define $\hat{r}' = \tau(\theta_{-j}, \theta'_j)$. Since τ is monotone and symmetric, it should be that $\hat{r}'_{jo} > \hat{r}'_{io}$. Hence, the DA implementation at problem $(P, \hat{r}')$ can never assign o to i instead of j. Since, φ does not have a sequential clinching implementation at problem θ, by Theorem 1, we have that $\#\varphi^\theta > 1$.

We now proceed with the proof of Theorem 2. We first prove the 'if' part. Suppose that for any outcome $\omega \neq \varphi(\theta)$, either ω is not stable at $(P, \tau(\theta))$, or it is stable at $(P, \tau(\theta))$, and there are two individuals i and j, such that

- i and j prefer each other's objects at ω more than their own ones, that is, $\omega(j) P_i \omega(i)$ and $\omega(i) P_j \omega(j)$,
- for any $o \notin \{\omega(i), \omega(j)\}$, $o P_i \omega(i)$ or $o P_j \omega(j)$.

Consider an arbitrary deviation $\omega \neq \varphi(\theta)$ and suppose that it is not stable at problem $(P, \hat{r})$. Then, by definition of stability, there are two individuals $i, j \in \mathcal{I}$ and object $o \in \mathcal{O}$ such that (1) $o\ P_i\ \omega(i)$, (2) $\omega(j) = o$, and (3) $\hat{r}_{io} > \hat{r}_{jo}$. However, since φ is DA-representable, there are no type reports $\theta_{-\{i,j\}}$ of individuals in $\mathcal{I} \setminus \{i, j\}$ for which φ produces such an outcome. Hence, i and j detect the deviation.

Now suppose there $\omega \neq \varphi(\theta)$ is stable at $(P, \hat{r})$, and it satisfies two preference conditions in the statement of Theorem 2. Namely, there are individuals i and j that prefer each others' objects at ω to their own, and there is no object $o \notin \{\omega(i), \omega(j)\}$ that is less preferred by both individuals than their respective objects at ω.

Suppose, for the sake of contradiction, that there are type reports $\theta'_{-\{i,j\}}$ of individuals in $\mathcal{I} \setminus \{i,j\}$, such that $\varphi(\theta_{\{i,j\}}, \theta'_{-\{i,j\}}) = \omega$. Let $\hat{r}' = \tau(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$. Since φ is DA-representable, ω should be the outcome of the DA implementation at the modified problem $(P_i, P_j, P'_{-\{i,j\}}, \hat{r}')$.

Consider the following mutually exclusive and collectively exhaustive scenarios: (i) j is assigned to $\omega(j)$ before i is assigned to $\omega(i)$, (ii) i is assigned to $\omega(i)$ before j is assigned to $\omega(j)$, (iii) i and j are assigned to their respective objects at ω in the same step. We study each of these cases separately.

Case 1. Suppose j is assigned to $\omega(j)$ before i is assigned to $\omega(i)$. Let t be the step of the DA implementation at the modified problem, when i claims and is assigned to $\omega(i)$, and j has already been tentatively assigned to $\omega(j)$. Since j prefers $\omega(i)$ over $\omega(j)$, it means that at the beginning of step t some individual in $\mathcal{I} \setminus \{i,j\}$ is tentatively assigned to $\omega(i)$.

Therefore, at the beginning of step t we have that no object is tentatively assigned to i, and both objects $\{\omega(i), \omega(j)\}$ are tentatively assigned to some individuals. Since $|\mathcal{I}| = |\mathcal{O}|$, there should be some object $o \notin \{\omega(i), \omega(j)\}$ that is not tentatively assigned to anyone. Since no object is simultaneously less preferred by both i and j than their own objects at ω, then either i prefers o to $\omega(i)$, or j prefers o more than $\omega(j)$. Either way, we arrive at a contradiction: i cannot prefer o more than $\omega(i)$ because she is claiming $\omega(i)$ at step t when o is available, and j cannot prefer o more than $\omega(j)$, because $\omega(j)$ is tentatively assigned to j, despite that o is not tentatively assigned to anyone (and hence, has not been claimed by anyone in any previous step).

Case 2. Suppose i is assigned to $\omega(i)$ before j is assigned to $\omega(j)$. This case is symmetric to Case 1, and hence, the proof is similar (identical).

Case 3. i and j are assigned to their respective objects at ω in the same step. The proof is similar to Case 1. Let t be the step of the DA implementation, when i and j claim and are assigned to their respective objects at ω. Since i prefers $\omega(j)$ over $\omega(i)$

and j prefers $\omega(i)$ over $\omega(j)$, it should be that both objects are tentatively assigned to some individuals in $\mathcal{I} \setminus \{i, j\}$. Therefore, at the beginning of step t, we have that no object is tentatively assigned to i and j, and both objects $\{\omega(i), \omega(j)\}$ are tentatively assigned to some individuals other than i and j. Since $|\mathcal{I}| = |\mathcal{O}|$, there should be some object $o \notin \{\omega(i), \omega(j)\}$ that is not tentatively assigned to anyone (in fact, there are at least two such objects). Since no object is simultaneously less preferred by both i and j than their own objects at ω, then either i prefers o to $\omega(i)$, or j prefers o more than $\omega(j)$. Either way, we arrive at a contradiction: both i and j are claiming their respective objects at ω in step t, despite that o is not tentatively assigned to anyone (and hence, has not been claimed by anyone in any previous step).

We now prove the 'only if' part. The proof is by contraposition. Suppose there is an outcome $\omega \neq \varphi(\theta)$ which is stable at $(P, \hat{r})$, and for any pair of individuals, one of the conditions in the statement of Theorem 3 fails. We will show that no two individuals detect the deviation ω, which would imply that $\#\varphi^\theta > 2$.

Consider arbitrary two individuals i and j. By the supposition above, i and j do not satisfy one of the preference conditions in the statement of Theorem 2. Namely, either one of the individuals in $\{i, j\}$ does not prefer the other individual's object at ω to her own, or both individuals in $\{i, j\}$ prefer the other individual's object at ω to their own, and there is an object $o \notin \{\omega(i), \omega(j)\}$ such that $\omega(i) P_i o$ and $\omega(j) P_j o$. We will prove $\#\varphi^\theta > 2$ for each of these cases separately.

Case 1. Suppose that one of the individuals in $\{i, j\}$ does not prefer the other individual's object at ω to her own. Without loss of generality, suppose $\omega(i) P_i \omega(j)$.

Consider the type reports $\theta'_{-\{i,j\}} = (P'_{-\{i,j\}}, r'_{-\{i,j\}})$ of individuals in $\mathcal{I} \setminus \{i, j\}$ where every individual in this set ranks her assigned object at ω as her first choice. Otherwise, the preferences and priorities agree with those in problem θ. To establish that i and j do not detect the deviation ω, it is sufficient to show that $\varphi(\theta_{\{i,j\}}, \theta'_{-\{i,j\}}) = \omega$.

Let $\hat{r}' = \tau(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$ denote the modified priorities at problem $(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$.

Since ω is a stable outcome at problem $(P, \hat{r})$, it should be that any object that i prefers to $\omega(i)$ is assigned to an individual with a higher $\hat{r}$ priority score than i at outcome ω. Similarly, any object that j prefers to $\omega(j)$ is assigned to an individual with a higher $\hat{r}$ priority score than j at outcome ω. Since τ is monotone, it should be that these priority scores' comparisons are preserved for $\hat{r}'$. Hence, during the implementation of DA at problem $(P_{\{i,j\}}, P'_{-\{i,j\}}, \hat{r}')$, individuals i and j will be rejected by all objects that they prefer more than $\omega(i)$ and $\omega(j)$, respectively. Hence, the DA implementation at problem $(P_{\{i,j\}}, P'_{-\{i,j\}}, \hat{r}')$ produces ω. Or equivalently, $\varphi(\theta_{\{i,j\}}, \theta'_{-\{i,j\}}) = \omega$.

Case 2. Suppose that both i and j prefer the other individual's assignment at ω to their own, and there is an object $o \notin \{\omega(i), \omega(j)\}$ such that $\omega(i) P_i o$, and $\omega(j) P_j o$.

Since $\omega(j) P_i \omega(i)$, $\omega(i) P_j \omega(j)$, and ω is stable at problem $(P, \hat{r})$, it should be that i has a higher $\hat{r}$ priority at $\omega(i)$ than j, and j has a higher $\hat{r}$ priority at $\omega(j)$ than i.

Let k be the individual with $\omega(k) = o$, where o is as defined above. Consider type reports $\theta'_{-\{i,j\}} = (P'_{-\{i,j\}}, r_{-\{i,j\}})$ of individuals in $\mathcal{I} \setminus \{i,j\}$, such that all individuals in this set other than k rank their assigned objects at ω as their first choice, and otherwise, the preference rankings and priority scores agree with $\theta_{-\{i,j\}}$. We construct the type report $\theta'_k = (P'_k, r'_k)$ of k as follows.

- k shares the same ranking of objects as i until (and including) $\omega(i)$, but she ranks o as her next choice right after $\omega(i)$,
- r'_k satisfies $r'_{k\tilde{o}} < r_{i\tilde{o}}$ for all o' with $o' P_i \omega(i)$ (or equivalently, for all $o' P'_k \omega(i)$.
- If $r_{i\omega(i)} > r_{j\omega(i)}$, then $r'_{k\omega(i)}$ is an arbitrary number in $(r_{j\omega(i)}, r_{i\omega(i)})$. Otherwise, $r'_{k\omega(i)}$ is an arbitrary number strictly smaller than $r_{i\omega(i)}$.

Let $\hat{r}' = \tau(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$ denote the modified priorities at problem $\theta_{\{i,j\}}, \theta'_{-\{i,j\}}$.

Claim. $\hat{r}'_{k\omega(i)} < \hat{r}'_{i\omega(i)}$ and $\hat{r}'_{k\omega(i)} > \hat{r}'_{j\omega(i)}$.

Proof. The first inequality in the claim directly follows from the construction of θ'_k and the equal treatment property of τ applied to i and k at problem $(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$.

We now prove $\hat{r}'_{k\omega(i)} > \hat{r}'_{j\omega(i)}$. The proof is by contradiction. Suppose $\hat{r}'_{k\omega(i)} < \hat{r}'_{j\omega(i)}$. Consider an alternative problem $\theta'' = (P'', r'')$ that differs from $(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$ by only that k's priority score $r''_{k\omega(i)}$ at object $\omega(i)$ is as follows. If $r_{i\omega(i)} > r_{j\omega(i)}$, then $r''_{k\omega(i)}$ is an arbitrary number strictly greater than $r_{i\omega(i)}$. Otherwise, $r''_{k\omega(i)}$ is an arbitrary number in $(r_{i\omega(i)}, r_{j\omega(i)})$. Let $\hat{r}'' = \tau(\theta'')$.

When moving from problem $(\theta_{\{i,j\}}, \theta'_{-\{i,j\}})$ to problem θ'', both the preference rankings of j and k, and the priority score comparison between j and k are unchanged. Hence, by the independence of irrelevant alternatives property of τ, and by our supposition, we have that $\hat{r}''_{k\omega(i)} < \hat{r}''_{j\omega(i)}$. Note that the type reports of i and j are unchanged across problems θ and θ''. Therefore, by the monotonicity property of τ, and by that $\hat{r}_{j\omega(i)} < \hat{r}_{i\omega(i)}$ (which we established in the beginning of the proof of Case 2 of the 'only if' part), we get $\hat{r}''_{j\omega(i)} < \hat{r}''_{i\omega(i)}$. Finally, by the equal treatment property of τ applied to i and k at problem θ'', we get $\hat{r}''_{i\omega(i)} < \hat{r}''_{k\omega(i)}$. Combing three inequalities above, we have,

$$\hat{r}''_{k\omega(i)} < \hat{r}''_{j\omega(i)} < \hat{r}''_{i\omega(i)} < \hat{r}''_{k\omega(i)},$$

a contradiction. $\square$

We now complete the proof of Case 2 of the 'only if' part.

Since ω is a stable outcome at $(P, \hat{r})$, it should be that any object that i prefers to $\omega(i)$ is assigned to an individual with a higher $\hat{r}$ priority than i at outcome ω. By the monotonicity property of τ is a monotone mapping, this is also true for $\hat{r}'$ priority scores. Similarly, any object that j that prefers to $\omega(j)$ is assigned to an individual with a higher $\hat{r}$ priority than j at outcome ω, and this is also true for $\hat{r}'$ priority scores.

Since k has a lower $\hat{r}'$ priority scores than i at all objects weakly more preferred than $\omega(i)$ (according to P'_k), during the implementation of DA at the modified problem $(P_{\{i,j\}}, P_{-\{i,j\}}, \hat{r}')$, she will be rejected by all objects strictly more preferred than $\omega(i)$, and she will be tentatively assigned at $\omega(i)$ at some step. Since $\hat{r}'_{k\omega(i)} > \hat{r}'_{j\omega(i)}$ (by the second part of the Claim), j will be eventually rejected by $\omega(i)$.

Continuing the DA implementation, i and j will be eventually rejected by all objects that they prefer strictly more than $\omega(i)$ and $\omega(j)$, respectively. Since $\hat{r}'_{k\omega(i)} > \hat{r}'_{i\omega(i)}$ (by the first part of the Claim), i will be tentatively assigned to $\omega(i)$, and k will be rejected by that object, and she will be assigned to o. In conclusion, the DA implementation at the modified problem $(P_{\{i,j\}}, P_{-\{i,j\}}, \hat{r}')$ would yield the outcome ω. Hence, $\varphi(\theta_{\{i,j\}}, \theta'_{-\{i,j\}}) = \omega$, and i and j do not detect the deviation.

Since i and j were arbitrary, we conclude that $\#\varphi^{\theta} > 2$. This complete the proof of Theorem 2.

A.4 Proof of Theorem 3

Before we start with the proof of Theorem 3, we introduce the following definitions: Given any sequential dictatorship φ, any θ and any $i \in \mathcal{I}$, let $n(\theta, i) \in \{1, \ldots, N\}$ be such that $i_{n(\theta,i)}(\theta) = i$. Given any $I' \subseteq \mathcal{I}$, let $S^{\varphi}(I') = \{\sigma \in S^{\varphi} | I' \subseteq \bar{I}(\sigma)$ and $\pi(\sigma) \in I'\}$.

We first prove the 'only if' part of the statement. Let φ be a sequential dictatorship that is not a vice dictatorship. We show that there exists a deviation that cannot be detected by any pair of individuals. We need to distinguish the following cases:

Case 1. Let φ be such that Definition 5 (3) is satisfied. We now further distinguish between the two subcases, where Definition 5 (2) is either satisfied or not.

Case 1.1 Suppose that Definition 5 (2) is satisfied and note that this implies that Definition 5 (1) is violated.

We now proceed in two steps: First, we establish a lemma that will be helpful to identify the problem and the individuals that will be affected by the deviation. Second, we construct the problem and the deviation and show that no pair of individuals can detect it.

Step 1:

We start with the lemma mentioned above.

Lemma 1. *There exists I' with $I' \cap (I_2^{\varphi} \cup I_3^{\varphi}) = \emptyset$ and $|I'| \geq 3$ such that (1) there* *are* $\hat{\sigma}^1, \hat{\sigma}^2 \in S^{\varphi}(I')$ *with* $\pi(\hat{\sigma}^1) \neq \pi(\hat{\sigma}^2)$ *and (2)* $|\bar{O}(\hat{\sigma}^1) \cap \bar{O}(\hat{\sigma}^2)| \geq 2$*.*

Proof. In the first part, we show that there is I' that satisfies condition (1). In the second part, we show that such I' automatically satisfies condition (2).

Part 1: We distinguish two cases:

First, suppose that the last two dictators on φ are always the same (i.e. for any pair of problems θ', θ'', we have $\bar{I}(\sigma_{N-2}(\theta')) = \bar{I}(\sigma_{N-2}(\theta''))$. Thus, since Definition 5 (2) is satisfied and Definition 5 (1) violated, there have to be two individuals $i', j' \notin I_2^{\varphi} \cup I_3^{\varphi}$ and two suboutcomes $\hat{\sigma}^1, \hat{\sigma}^2 \in S^{\varphi}$, where $\pi(\hat{\sigma}^1) = i'$, $\pi(\hat{\sigma}^2) = j'$, $|\bar{I}(\hat{\sigma}^1)| \geq 3$ and $|\bar{I}(\hat{\sigma}^2)| \geq 3$. Select an arbitrary $k' \in I_N^{\varphi}$. Hence, $k' \notin I_n^{\varphi}$ for all $n \leq N-2$. Then, the conditions in (1) are satisfied for $I' = \{i', j', k'\}$.

Second, assume that the last two dictators on φ are not always the same. Thus, there are two problems θ', θ'' such that there is an individual $i' \in \mathcal{I}$, where $i' \in \bar{I}(\sigma_{N-2}(\theta'))$ and $i' \notin \bar{I}(\sigma_{N-2}(\theta''))$. Hence, for $I' = \{i'\} \cup \bar{I}(\sigma_{N-2}(\theta''))$, we have $\sigma_{N-2}(\theta'') \in S^{\varphi}(I')$, $\pi(\sigma_{N-2}(\theta'')) = i'$ and $|\bar{I}(\sigma_{N-2}(\theta''))| \geq 3$. Let $j', k' \in \bar{I}(\sigma_{N-2}(\theta''))$ such that $n(\theta', j') < n(\theta', k')$. Thus, $n(\theta', j') < n(\theta', i')$. This ensures that $\sigma_{n(\theta',j')}(\theta') \in S^{\varphi}(I')$ with $\pi(\sigma_{n(\theta',j')}(\theta')) = j'$. Then, the conditions in (1) are satisfied for $I' = \{i', j', k'\}$. *Part 2:* We now show that any I' that satisfies condition (1) also satisfies condition (2). Concretely, consider I' that satisfies condition (1).

We proceed by contradiction: Suppose that for any pair $\sigma, \sigma' \in S^{\varphi}(I')$, with $\pi(\sigma) \neq \pi(\sigma')$, we have $|\bar{O}(\sigma) \cap \bar{O}(\sigma')| < 2$. By condition (1) there are $\hat{\sigma}^1, \hat{\sigma}^2 \in S^{\varphi}(I')$ with $\pi(\hat{\sigma}^1) = i'$, $\pi(\hat{\sigma}^2) = j'$ and $i' \neq j'$. Assume that $|\bar{O}(\hat{\sigma}^1) \cap \bar{O}(\hat{\sigma}^2)| = 0$. Thus, there exist two objects o^*, o^{**} both in $\bar{O}(\hat{\sigma}^1)$ and both not in $\bar{O}(\hat{\sigma}^2)$ and two objects $\tilde{o}, \tilde{o}'$ both in $\bar{O}(\hat{\sigma}^2)$ and both not in $\bar{O}(\hat{\sigma}^1)$.

Next, consider a problem θ^* where for all $\bar{k} \in \mathcal{I}$, for each $\hat{o}^1 \in \mathcal{O} \setminus \{\tilde{o}, \tilde{o}', o^*\}$ and for each $\hat{o}^2 \in \{\tilde{o}, \tilde{o}', o^*\}$, we have $\hat{o}^1 \ \theta^*_{\bar{k}} \ \hat{o}^2$. Thus, there exists $\hat{\sigma} \in S^{\varphi}(I)$, where $\{\tilde{o}, \tilde{o}', o^*\} \subseteq \bar{O}(\hat{\sigma})$. Hence, $\pi(\hat{\sigma}) = i'$. Consider θ^{**} such that for all $\bar{k} \in \mathcal{I}$, for each $\hat{o}^3 \in \mathcal{O} \setminus \{\tilde{o}, o^*, o^{**}\}$ and for each $\hat{o}^4 \in \{\tilde{o}, o^*, o^{**}\}$, we have $\hat{o}^3 \ \theta^{**}_{\bar{k}} \ \hat{o}^4$. This implies that there exists $\bar{\sigma} \in S^{\varphi}(I')$, where $\{\tilde{o}, o^*, o^{**}\} \subseteq \bar{O}(\bar{\sigma})$. Hence, $\pi(\bar{\sigma}) = i'$. However, note that $|\bar{O}(\hat{\sigma}^2) \cap \bar{O}(\bar{\sigma})| \geq 2$. Thus, since $\pi(\hat{\sigma}^2) = j'$ we arrive at a contradiction. A similar argument will apply to the case where $|\bar{O}(\hat{\sigma}^1) \cap \bar{O}(\hat{\sigma}^2)| = 1$ and is therefore omitted.

This completes the proof of the lemma. $\square$

We proceed as follows: we choose an arbitrary set of individuals $I \subseteq \mathcal{I}$ that satisfies the conditions of Lemma 1. This set will contain the individuals that are affected by the deviation we are about to construct in Step 2 below.

Specifically, by Lemma 1, there is $\sigma^1, \sigma^2 \in S^{\varphi}(I)$ such that $\pi(\sigma^1) \neq \pi(\sigma^2)$ and for which $|\bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)| \geq 2$. Denote $\pi(\sigma^1) = i$, $\pi(\sigma^2) = j$ and consider an arbitrary pair of objects $o', o'' \in \bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)$. Now note that by definition of S^{φ} there exist $\hat{\theta}$ and $\tilde{\theta}$ such that there are $n, n' \in \{1, \ldots, N-1\}$ with $\sigma_n(\hat{\theta}) = \sigma^1$ and $\sigma_{n'}(\tilde{\theta}) = \sigma^2$. Also note that for every $\hat{\theta}^*$ such that for all $i^* \notin \bar{I}(\sigma^1)$, $\varphi(\hat{\theta})(i^*)$ is ranked first on $\hat{\theta}^*_{i^*}$, we must reach $\sigma_n(\hat{\theta}^*) = \sigma^1$.

Step 2:

Construction of the problem: Consider the following problem θ that will be the basis for the deviation: First, for all $k^* \notin \{i, j\}$, object $\varphi(\hat{\theta})(k^*)$ ranks first on θ_{k^*} and for each $\bar{o} \notin \{o', o''\}$, let $\bar{o}\ \theta_{k^*}\ o'$ and $\bar{o}\ \theta_{k^*}\ o''$. Second, suppose that θ_i and θ_j rank o' first and o'' second.

It is easy to check that $\varphi(\theta)(i) = o'$, $\varphi(\theta)(j) = o''$ and $\varphi(\theta)(k^*) = \varphi(\hat{\theta})(k^*)$ for all $k^* \notin \{i, j\}$.

Construction of the deviation: We are now ready to construct the deviation $\omega \neq \varphi(\theta)$ at problem θ:

- For all $k^* \in \mathcal{I} \setminus \{i, j\}$, let $\omega(k^*) = \varphi(\theta)(k^*)$.
- $\omega(i) = o''$ and $\omega(j) = o'$.

Detectability: Take any $\hat{I} \subset \mathcal{I}$ with $|\hat{I}| = 2$. We verify that $\hat{I}$ does not detect ω considering the following exhaustive case distinction.

Case 1.1.1: Suppose that $\hat{I} \cap \{i, j\} = \emptyset$. Since for all $k^* \in \mathcal{I} \setminus \{i, j\}$ we have $\omega(k^*) = \varphi(\theta)(k^*)$, we know that $\hat{I}$ does not detect ω.

Case 1.1.2: Suppose that $\hat{I} = \{i, j\}$. First, note that for every $\tilde{\theta}^*$ where for all $j^* \notin \bar{I}(\sigma^2)$, object $\varphi(\tilde{\theta})(j^*)$ ranks first on $\tilde{\theta}^*_{j^*}$, we must reach $\sigma^2 = \sigma_{n'}(\tilde{\theta}^*)$. Now, consider problem θ' as follows: For all $j^* \notin \bar{I}(\sigma^2)$ specify that $\varphi(\tilde{\theta})(j^*)$ ranks first on θ_{j^*} and for each $\bar{o} \notin \{o', o''\}$, let $\bar{o}\ \theta_{j^*}\ o'$ and $\bar{o}\ \theta_{j^*}\ o''$. Furthermore, suppose that $\theta'_i = \theta_i$ and $\theta'_j = \theta_j$ and that for all $\bar{j} \in \bar{I}(\sigma^2) \setminus \{i, j\}$, objects o', o'' rank last on $\theta'_{\bar{j}}$. Calculating the outcome reveals that $\omega(i) = \varphi(i)(\theta')$ and $\omega(j) = \varphi(j)(\theta')$. Thus, $\hat{I}$ does not detect ω.

Case 1.1.3. Suppose that $|\hat{I} \cap \{i, j\}| = 1$.

- If $\hat{I} = \{i_d, i\}$, then consider $\bar{\theta}$ such that for all $l^* \notin \hat{I}$, o' is ranked first on $\bar{\theta}_{l^*}$ and o'' ranks last on $\bar{\theta}_{l^*}$. Let $\bar{\theta}_i = \bar{\theta}_i$ and $\bar{\theta}_{i_d} = \theta_{i_d}$. Thus, $\omega(i_d) = \varphi(\bar{\theta})(i_d)$. This implies that $n(\bar{\theta}, i) \geq 3$. Hence, $\omega(i) = \varphi(\bar{\theta})(i) = o''$.

- If $\hat{I} = \{i, \hat{j}\}$ with $\hat{j} \neq i_d$, then consider $\bar{\theta}$ such that for all $l^* \notin \hat{I}$, o' is ranked first on $\bar{\theta}_{l^*}$ and $\omega(\hat{j})$ ranks last on $\bar{\theta}_{l^*}$. Let $\bar{\theta}_i = \theta_i$ and $\bar{\theta}_{\hat{j}} = \theta_{\hat{j}}$. Thus, $\omega(i) = \varphi(\bar{\theta})(i)$ and $\omega(\hat{j}) = \varphi(\bar{\theta})(\hat{j})$.

- If $\hat{I} = \{j, \hat{j}\}$, then consider $\bar{\theta}$ such that for all $l^* \notin \hat{I}$ and for each $\bar{o} \notin \{o'', \omega(\hat{j})\}$, we have $\bar{o}\ \theta_{l^*}\ o''$ and $\bar{o}\ \theta_{l^*}\ \omega(\hat{j})$. Let $\bar{\theta}_i = \theta_i$ and $\bar{\theta}_{\hat{j}} = \theta_{\hat{j}}$. Thus, $\omega(j) = \varphi(\bar{\theta})(j) = o'$ and $\omega(\hat{j}) = \varphi(\bar{\theta})(\hat{j})$.

Hence, in all three instances above, $\hat{I}$ does not detect ω.

Since the selection of $\hat{I}$ was arbitrary, we can conclude $\#\varphi > 2$. This completes the arguments for this case.

Case 1.2. Suppose that φ violates Definition 5 (2). We proceed in a similar manner as in Case 1.1. We first identify the individuals that will be affected by the deviation we aim to construct. In a second step, we will apply the arguments in Step 2 of Case 1.1.

Case 1.2.1. Suppose that I_2^φ is not a singleton. We first show that there exists $I = \{i, j\}$ with $I \subseteq I_2^\varphi \cup I_3^\varphi$ such that there exist $\sigma^1, \sigma^2 \in S^\varphi(I)$, where $\pi(\sigma^1) = i$, $\pi(\sigma^2) = j$, $|\bar{O}(\sigma^1)| = N - 2$ and $|\bar{O}(\sigma^2)| = N - 1$. We need to consider the following subcases:

Case 1.2.1.1. Suppose that $I_2^\varphi = I_3^\varphi$. Then, since Definition 5 (2) is violated, we must have $|I_2^\varphi| = |I_3^\varphi| \geq 3$. Thus, there exists $j \in I_2^\varphi$ and $\hat{\theta}$, where $n(\hat{\theta}, j) \geq 4$. Let $i \in I_3^\varphi$ such that $n(\hat{\theta}, i) = 3$. Since $j \in I_2^\varphi$, there exists $\tilde{\theta}$ such that $n(\tilde{\theta}, j) = 2 > n(\tilde{\theta}, i)$. Thus, we can select $I = \{i, j\}$, $\sigma^1 = \sigma_2(\hat{\theta})$ and $\sigma^2 = \sigma_1(\tilde{\theta})$ to satisfy the claim above.

Case 1.2.1.2. Suppose that $I_2^\varphi \neq I_3^\varphi$ and that there exists $i' \in I_2^\varphi$ with $i' \notin I_3^\varphi$. Thus, there exists $j \in I_2^\varphi$ and $\hat{\theta}$, where $n(\hat{\theta}, j) \geq 4$. Let $i \in I_3^\varphi$ be such that $n(\hat{\theta}, i) = 3$. Since $j \in I_2^\varphi$, there exists $\tilde{\theta}$ such that $n(\tilde{\theta}, j) = 2 > n(\tilde{\theta}, i)$. Thus, we can select $I = \{i, j\}$, $\sigma^1 = \sigma_2(\hat{\theta})$ and $\sigma^2 = \sigma_1(\tilde{\theta})$ to satisfy the claim above.

Case 1.2.1.3. Suppose that $I_2^\varphi \neq I_3^\varphi$ and that there exists $i' \in I_3^\varphi$ with $i' \notin I_2^\varphi$. Thus, we can select $i \in I_3^\varphi$ and $\hat{\theta}$ such that $n(\hat{\theta}, i) = 3$ and $i \notin I_2^\varphi$. Moreover, because I_2^φ is not a singleton, we can select $j \in I_2^\varphi$ with $n(\hat{\theta}, j) \geq 4$. Also, since $j \in I_2^\varphi$, there exists $\tilde{\theta}$ such that $n(\tilde{\theta}, j) = 2 > n(\tilde{\theta}, i)$. Thus, we can select $I = \{i, j\}$, $\sigma^1 = \sigma_2(\hat{\theta})$, $\sigma^2 = \sigma_1(\tilde{\theta})$ to satisfy the claim above.

For each of the three subcases above, since $N \geq 5$ and a single object is assigned in σ^2 and only two objects are assigned in σ^1, we must have $|\bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)| \geq 2$. Consider an arbitrary pair of objects in $o', o'' \in \bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)$. Also note that for every $\hat{\theta}^*$, where for all $i^* \notin \bar{I}(\sigma^1)$, object $\varphi(\hat{\theta})(i^*)$ is ranked first on $\hat{\theta}^*_{i^*}$, we eventually must reach $\sigma_2(\hat{\theta}^*) = \sigma^1$. We now can apply the arguments of Step 2 in Case 1.1.

Case 1.2.1. Suppose that I_2^φ is a singleton. Thus, $|I_3^\varphi| \geq 2$ and $I_2^\varphi \cap I_3^\varphi = \emptyset$. Note that since I_2^φ is a singleton and $|I_3^\varphi| \geq 2$, there must exist $\sigma^1, \sigma^2 \in S^\varphi$ such that $|\bar{O}(\sigma^1)| = N - 2$ and $|\bar{O}(\sigma^2)| = N - 2$ as follows: there is a unique object o^* that is assigned in both σ^1 and σ^2 (that is the other assigned objects and the two suboutcomes are not the same) and for which $\pi(\sigma^1) \neq \pi(\sigma^2)$. Otherwise, we would reach a contradiction to $|I_3^\varphi| \geq 2$. Now note that $N \geq 5$ implies that there is a pair of objects $\tilde{o}, \hat{o} \in \bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)$.

We now consider $I = \{i, j\}$ such that $\pi(\sigma^1) = i$ and $\pi(\sigma^2) = j$ and we choose an arbitrary pair $o', o'' \in \bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)$. Again, note that for every $\hat{\theta}^*$ such that for all $i^* \notin \bar{I}(\sigma^1)$, $\varphi(\hat{\theta})(i^*)$ is ranked first on $\hat{\theta}^*_{i^*}$, we must reach $\sigma_n(\hat{\theta}^*) = \sigma^1$. We now can apply the arguments of Step 2 in Case 1.1. and thereby complete the arguments for this subcase.

This completes the arguments for Case 1.2.

Case 2. Let φ be such that Definition 5 (3) is violated. We proceed in a similar manner as in the previous cases.

Since Definition 5 (3) is violated, there must exist a pair $\sigma^1, \sigma^2 \in S^\varphi$, with $\bar{O}(\sigma^1) = \bar{O}(\sigma^2)$ and $\bar{I}(\sigma^1) = \bar{I}(\sigma^2)$ such that $\pi(\sigma^1) \neq \pi(\sigma^2)$. Denote $\pi(\sigma^1) = i$, $\pi(\sigma^2) = j$ and let $\{o', o''\} \subseteq \bar{O}(\sigma^1) \cap \bar{O}(\sigma^2)$. Now, consider two problems $\hat{\theta}, \tilde{\theta}$ such that $\sigma^1 = \sigma_n(\hat{\theta})$ and $\sigma^2 = \sigma_n(\tilde{\theta})$ for some step n. Again, for every $\hat{\theta}^*$ such that for all $i^* \notin \bar{I}(\sigma^1)$, object $\varphi(\hat{\theta})(i^*)$ is ranked first on $\hat{\theta}^*_{i^*}$, we must reach $\sigma_n(\hat{\theta}^*) = \sigma^1$. Finally, apply Step 2 of Case 1.1.

This completes the arguments for Case 2, and hence, the 'only if' direction.

We proceed with the 'if' part of the statement. Suppose that φ is a vice dictatorship. To establish that $\#\varphi = 2$, consider an arbitrary problem θ and an arbitrary deviation $\omega \neq \varphi(\theta)$. We have to show that there is a group of two individuals that can detect the deviation. Consider the smallest step $n \in \{1, \dots N-1\}$ such that $\omega(i_n(\theta)) \neq \varphi(\theta)(i_n(\theta))$. Thus, there exists $n' > n$ such that $\omega(i_{n'}(\theta)) \neq \varphi(\theta)(i_{n'}(\theta))$, where $\omega(i_{n'}(\theta)) = \varphi(\theta)(i_n(\theta))$. Denote $i_n(\theta) := i$ and $i_{n'}(\theta) := j$.

Case 1: Suppose that $\{i, j\} = I_2^\varphi \cup I_3^\varphi$. Since $\omega(i_d) = \varphi(\theta)(i_d)$, we have $i = i_2(\theta')$ for all θ' with $\theta'_i = \theta_i$, where $\varphi(\theta')(i_d) = \varphi(\theta)(i_d)$ and thus $\varphi(\theta')(i) = \varphi(\theta)(i) \neq \omega(i)$. Hence, $\hat{I} = \{i, i_d\}$ detect ω.

Case 2: Suppose that $\{i, j\} = I_{N-1}^\varphi \cup I_N^\varphi$. Then $\{\omega(i), \omega(j)\} = \bar{O}(\sigma_{N-2}(\theta))$, since otherwise there must exist $k \notin \bar{I}(\sigma_{N-2}(\theta))$ such that $\omega(k) \neq \varphi(\theta)(k)$, which would contradict the assumption that n is the smallest step for which $\omega(i_n(\theta)) \neq \varphi(\theta)(i_n(\theta))$. Also notice that $\{\omega(i), \omega(j)\} = \bar{O}(\sigma_{N-2}(\theta))$ implies that $\omega(i) = \varphi(\theta)(j)$ and $\omega(j) = \varphi(\theta)(i)$. Hence, $\omega(j) = \varphi(\theta)(i)\ \theta_i\ \omega(i)$. Moreover, by Definition 5 (3), for any pair

$\sigma, \hat{\sigma} \in S^{\varphi}$ with $\bar{O}(\sigma) = \bar{O}(\hat{\sigma})$ and $\bar{I}(\sigma) = \bar{I}(\hat{\sigma})$, we must have $\pi(\sigma) = \pi(\hat{\sigma})$. This implies that $\hat{I} = \{i, j\}$ detect ω.

Case 3: Finally, suppose that $\{i, j\} \neq I_2^{\varphi} \cup I_3^{\varphi}$ and $\{i, j\} \neq I_{N-1}^{\varphi} \cup I_N^{\varphi}$. If $i = i_d$, then i_d detects the deviation since for all θ' with $\theta'_i = \theta_i$, it must be $\varphi(\theta')(i_d) = \varphi(\theta)(i_d)$. For $i \neq i_d$, it must be $\omega(j) = \varphi(\theta)(i)\ \theta_i\ \omega(i)$. Since $\{i, j\} \neq I_2^{\varphi} \cup I_3^{\varphi}$ and $\{i, j\} \neq I_{N-1}^{\varphi} \cup I_N^{\varphi}$, we know that for all θ' with $\theta'_i = \theta_i$, it must be $\varphi(\theta')(i)\ \theta'_i\ \varphi(\theta')(j)$. Thus, $\hat{I} = \{i, j\}$ detect ω.

Since we selected the problem and deviation arbitrarily, we conclude that any deviation can be detected by at most two individuals. Thus, $\#\varphi = 2$. This completes the proof for Theorem 3.

A.5 Proof of Theorem 4

We first prove the 'if' part. Suppose φ is a dual-dictatorship. Consider an arbitrary problem $b \in \Theta$ and a deviation $\omega \neq \varphi(b)$. Let i be some individual for whom $\omega(i) \neq \varphi(b)(i)$. Also, let $(x, y) = \omega(i)$ and $(x', y') = \varphi(b)(i)$. Consider cases:

1. $x = x'$. We show that i detects the deviation ω. Suppose, for the sake of contradiction, that she does not detect the deviation, i.e., there is problem $\bar{b} \in \Theta$ with $\bar{b}_i = b_i$, such that $(x, y) = \varphi(\bar{b})(i)$. Since, i has the same bid and allocation at problems b and $\bar{b}$, and φ is a fixed-pay auction, we should have that $y = y'$. This contradicts that $(x, y) = (x', y')$.

2. $x \neq x'$. The condition means that i receives the object under one of the outcomes ω or $\varphi(b)$, and she does not receive the object under the other outcome. Let i_1 be the first dictator in the definition of the dual-dictatorship. Then, it should be that i_1 too receives the object under one of the outcomes ω or $\varphi(b)$, and she does not receive the object under the other outcome. However, by definition, whether i_1 receives an object or not is fully determined by her bid b_{i_1}. Hence, i_1 detects the deviation ω.

We proved that in both cases a single individual detects the deviation ω. Since $b \in \Theta$

and deviation $\omega \neq \varphi(b)$ were arbitrary, we conclude that $\#\varphi = 1$.

We now prove the 'only if' part. Suppose $\#\varphi = 1$.

Claim 1. φ is a fixed-pay auction.

Proof. We prove the claim by contradiction. Suppose, there are problems $b, b' \in \Theta$ and an individual i with $b_i = b'_i$, such that for $(x, y) := \varphi(b)(i)$ and $(x', y') := \varphi(b')(i)$, we have that $x = x'$ and $y \neq y'$. For the problem b, consider the deviation ω that differs from $\varphi(b)$ only by that $\varphi(b)(i) = (x', y')$. Then, no single individuals detects this deviation, contradicting $\#\varphi = 1$. □

Let $\bar{I} \subseteq \mathcal{I}$ be the set of individuals that receive the object under some problem. If $\bar{I}$ is singleton, then φ is a dual-dictatorship with $\bar{B} = \mathbb{R}_+$. For the rest of the proof, we assume that $\bar{I}$ is not a singleton.

Claim 2. There is an $i_1 \in \bar{I}$ and a $\bar{B} \subseteq \mathbb{R}_+$, such that for all $b \in \Theta$ and $(x, y) = \varphi(b)(i_1)$,

$$x = 1 \text{ if and only if } b_{i_1} \in \bar{B}.$$

Proof. We prove the claim by contradiction. Suppose, for all $i \in \bar{I}$ there is a bid $b_i \in \mathbb{R}_+$ and others' bids b'_{-i} and $\tilde{b}_{-i}$, with $(x', y') = \varphi(b_i, b'_{-i})$ and $(\tilde{x}, \tilde{y}) = \varphi(b_i, \tilde{b}_{-i})$, such that $x = 1$ and $\tilde{x} = 0$. Fix the problem $b = (b_i)_{i \in \mathcal{I}}$, and let i be the individual who receives the object at $\varphi(b)$. Consider a deviation $\omega \neq \varphi(b)$ such that some individual $i' \in \bar{I} \setminus \{i\}$ receives the object. Then, no single individual detects the deviation ω. □

Claim 3. $|\bar{I}| = 2$.

Proof. We prove the claim by contradiction. Suppose $|\bar{I}| > 2$. For any $i \in \bar{I} \setminus \{i_1\}$, let the bid b_i be such that for some bids of others b'_{-i}, i receives the object at the outcome $\varphi(b_i, b'_{-i})$. Fix the bids of individuals in $\bar{I} \setminus \{i_1\}$ at $(b_i)_{i \in \bar{I} \setminus \{i_1\}}$ and set the bid of i_1 at some $b_{i_1} \notin \bar{B}$. Let b denote the corresponding problem (where the bids of individuals in $\mathcal{I} \setminus \bar{I}$ are arbitrary). Then, it should be that some individual in

$i \in \bar{I} \setminus \{i_1\}$ receives the object. Consider a deviation $\omega \neq \varphi(b)$ such that some other individual $i' \in \bar{I} \setminus \{i_1, i\}$ receives the object (such a i' exist since we assumed that $|\bar{I}| > 2$). Then, no single individual detects the deviation ω, which contradicts that $\#\varphi = 1$. $\square$

By Claims 2 and 3, i_1 and $i_2 := \bar{I} \setminus \{i_1\}$ are the dual-dictators of φ. This completes the proof of Theorem 4.

A.6 Proof of Theorem 5

The 'if' part is trivial: if a social choice mechanism is dictatorial, then any deviation is detected by the dictator (i.e., the individual $\bar{i}$ in the definition of the dictatorial social choice mechanism).

We now prove the 'only if' part. The proof is by contraposition. Suppose the social choice mechanism φ is not dictatorial. Then, there is no individual $i \in \mathcal{I}$, such that $\varphi(\theta) = \varphi(\theta')$ for any problems θ and θ' with $\theta_i = \theta'_i$. Equivalently, for any individual $i \in \mathcal{I}$, there are two problems θ and θ' with $\theta_i = \theta'_i := \bar{\theta}_i$, such that $\varphi(\theta) \neq \varphi(\theta')$. Consider the problem $\bar{\theta} = (\bar{\theta}_i)_{i \in \mathcal{I}}$, and the deviation $\omega \neq \varphi(\bar{\theta})$. It is immediate from the construction of $\bar{\theta}$ that no individual detects the deviation ω. Thus, $\#\varphi > 1$.

A.7 Proof of Proposition 8

The proof has four parts. In Part 1 we show that $\#RSF \leq R + 2$. In Part 2 we construct a problem θ such that $\#RSF^{\theta} \geq R + 2$. Parts 1 and 2 jointly establish that $\#RSF = R+2$. In Part 3 we construct a problem θ such that $\#OSF^{\theta} \geq R+2$. Finally, in Part 4 we construct a problem θ such that $\#OSF^{\theta} \geq Q - R + 1$. Parts 3 and 4 jointly establish that $\#OSF \geq \max\{Q - R + 1, R + 2\}$.

Part 1. $\#RSF \leq R + 2$.

First, we introduce several definitions and state a lemma about the RSF mechanism. For a given outcome $\omega \in \Omega$, and individuals $i, j \in \mathcal{I}$, we say that i's **priority is**

violated by j, if $\omega(i) = 0, \omega(j) = 1$, and $\rho_i > \rho_j$. We say an outcome ω is **within-type priority compatible**, if whenever i's priority is violated by j, it should be that

$$i \in I_H \text{ and } j \in I_L.$$

We say that an outcome ω is **saturated priority compatible**, if whenever i's priority is violated by j, $i \notin I_L$ and $j \in I_L$, then it should be that

$$\left|\left\{i' \in I_L : \omega(i') = 1\right\}\right| = R.$$

We say a choice mechanism φ is within-type priority compatible (or saturated priority compatible), if the outcome $\varphi(\theta)$ is within-type priority compatible (or saturated priority compatible) for all $\theta \in \Theta$.

The following result is a special case of Theorem 3 in Imamura (2020).

Lemma 2. *In the choice with affirmative action setting, RSF is the unique within-type priority compatible and saturated priority compatible mechanism.*

Now we prove $\#RSF \leq R + 2$. Consider an arbitrary problem θ and an arbitrary deviation $\omega \neq RSF(\theta)$. By Lemma 2, the outcome ω is either not within-type priority compatible, or not saturated priority compatible. We consider each case separately.

First, suppose ω is not within-type priority compatible. Then, there are individuals i and j, such that i's priority is violated by j, and one of the following holds: (i). $i, j \in I_L$, (ii). $i, j \in I_H$, or (iii). $i \in I_L$ and $j \in I_H$. None of these cases can happen under the RSF mechanism, hence, $\{i, j\}$ detects the deviation ω.

Now, suppose ω is not saturated priority compatible. Then, there are individuals i and j, such that i's priority is violated by j, $i \in I_H, j \in I_L$, and

$$\left|\left\{i' \in I_L : \omega(i') = 1\right\}\right| > R.$$

Let $\tilde{I} \subseteq \left\{i' \in I_L(\theta) : \omega(i') = 1\right\}$ be the lowest priority $R+1$ individuals in the subset. It is immediate from the description of RSF that i should have been chosen over some individual in $\tilde{I}$. Hence, $\tilde{I} \cup \{i\}$ (whose size is $R + 2$) detects the deviation ω.

Since the problem θ and deviation ω was arbitrary, we conclude that $\#RSF \leq R+2$.

Part 2. There is problem θ, such that $\#RSF^{\theta} \geq R+2$.

Consider a problem θ, where all individuals in I_L have lower priorities than all individuals in I_H.

For this problem, RSF chooses R individuals with the highest priorities from I_L and $Q-R$ individuals with the highest priorities from I_H. Consider the deviation $\omega \neq RSF(\theta)$, that chooses all $R+1$ highest priority individuals in I_L and only $Q-R+1$ highest priority individuals from I_H. Then, no subset of individuals of size strictly less than $R+2$ detects the deviation ω. Hence, $\#OSF^{\theta} \geq R+2$.

Part 3. There is a problem θ, such that $\#OSF^{\theta} \geq R+2$.

The proof of this part is identical to that of Part 2. Namely, we use the same problem and deviation to prove the result.

Part 4. There is a problem θ, such that $\#OSF^{\theta} \geq Q-R+1$.

Consider a problem θ where $\min\{Q-R, |I_L|-R\}$ individuals in I_L have highest priorities among all individuals $\mathcal{I}$, and the remaining individuals in I_L have the lowest priorities among all individuals $\mathcal{I}$. For this problem, the OSF mechanism would chose $\min\{Q-R, |I_L|-R\}+R = \min\{Q, |I_L|\}$ highest priority individuals in I_L.

Let i be the highest priority non-chosen individual in $I \setminus I_L$ under $OSF(\theta)$, and let $\tilde{i}$ be the lowest priority chosen individual in I_L. Consider the deviation ω that differs from $OSF(\theta)$ by that $\omega(i)=1$ and $\omega(\tilde{i})=0$. Let $I \subseteq \mathcal{I}$ be an arbitrary subset of individuals with $|I| < Q-R+1$. Consider cases:

(i). Suppose $I \subseteq I_L$. Since the outcome of individuals in $I \setminus \{\tilde{i}\}$ is the same under $OSF(\theta)$ and ω, this subset does not detect the deviation.

(ii). Suppose $I \cap I_H \neq \emptyset$. Since $I_H \neq \emptyset$ and $|I| < Q-R+1$, we have that $|I \cap I_L| < Q-R$. Consider the problem θ', such that $\theta'_I = \theta_I$, and individuals in $\left(I \cap I_L \setminus \{\tilde{i}\}\right) \cup \{i\}$ have higher priorities than everyone in $\mathcal{I} \setminus I$. Then, the set of chosen individuals from I is the same under both ω and $OSF(\theta')$. Hence, I does not detect the deviation ω. This completes the proof of Proposition 8.

B Application-Rejection (AR) Mechanisms

The *application-rejection* mechanism with *permanency-execution period* e, denoted by AR_e, selects outcomes as follows. We follow the description in Chen and Kesten (2017). Take any problem θ:

Round $t = 0$. *Each individual claims her most preferred object. Each object that is claimed by some individuals, is assigned to the claimant with the highest priority and rejects the rest.*

(in general) Rejected individuals that have not yet claimed their e-th most preferred object, claim their most preferred objects that has not rejected them. In case an individual is rejected from her top e most preferred objects, she becomes passive in this round. Each object that is claimed by some individuals, is tentatively assigned to the claimant with the highest priority and rejects the rest. The round terminates either when each individual is assigned or passive. All tentatively assigned individuals become unavailable, and we proceed to the next round.

Round $t \geq 1$. *Each available individual claims her te+1-st most preferred object. Each object with claimants, is assigned to the one with the highest priority and rejects the rest.*

(in general) Rejected individuals that have not yet claimed their te+e-th most preferred objects, claim their most preferred objects that has not rejected them. In case an individual is rejected from her top $te + e$ most preferred objects, she remains passive in this round. Each object that is claimed by some individuals, is tentatively assigned to the claimant with the highest priority and rejects the rest. The round terminates either when each individual is either tentatively assigned or passive. All tentatively assigned individuals become unavailable, and we proceed to the next round.

The algorithm terminates when each agents has been assigned to an object. In this case, all assignments become final and the resulting outcome is $AR_e(\theta)$.

C An Application of the Sequential Clinching Characterization

In this section, we elaborate on the auditability properties of serial dictatorships which are formally introduced in section 4.3. We have already established that these mechanisms have a worst-case auditability index of two (Theorem 3) sinc each serial dictatorship is a vice dictatorship in the sense of Definition 5. The question is whether there are some problems for which serial dictatorships have an auditability index of one, instead of two. We show that this 'rarely' happens. We will use the sequential clinching characterization result for auditability index of one (Theorem 1) to establish this result. More specifically, we will show that for a 'generic' problem, a serial dictatorship mechanism does not have a sequential clinching implementation.

Consider an arbitrary serial dictatorship φ, and let us index individuals based on the fix dictatorial ordering. $\mathcal{I} = \{i_1, i_2, \ldots, i_N\}$. That is, $i_n = I_n^\varphi$ for all $n \in \{1, \ldots, N\}$.

Proposition 9. *As $N \to \infty$, the proportion of problems for which φ has an audiability index of one converges to zero.*

Proposition 9 follows from a lemma that we state and prove next.

In what follows we fix a problem θ. For each $n \in \{1, 2, \ldots, N\}$, let $\bar{O}_n$ denote the n most preferred objects of i_n.

Lemma 3. *A serial dictatorship φ has a sequential clinching implementation at problem θ if and only if $\bar{O}_n \subsetneq \bar{O}_{n+1}$, for all $n \in \{1, 2, \ldots, N-1\}$.*

Proof. Let $O_{i_n}(\theta_{i_n})$ be the set of possible objects for i_n at type report θ_{i_n}. From the description of the serial dictatorship, we can conclude that each i_n is guaranteed to be assigned to one of her n most preferred objects. Moreover, she may be assigned to any of the n most preferred objects, depending on the type reports of earlier dictators. Hence, for each $n \in \{1, 2, \ldots, N\}$, we have that

$$O_{i_n}(\theta_{i_n}) = \bar{O}_n.$$

We now prove the Lemma, and we start with the proof of the 'if' part. Suppose $\bar{O}_n \subsetneq \bar{O}_{n+1}$ for all $n \in \{1, 2, \ldots, N-1\}$. We will show that the serial dictatorial implementation at problem θ is the desired sequential clinching implementation of φ.

The proof is by induction. It is immediate that $i_1 = i_d$ (the first dictator) is clinching her most preferred object o_1. Now suppose the claim holds for up to some step $t \geq 1$. Consider the step $t+1$ of the serial dictatorial implementation. Since i_t clinches some object among available ones at step t (by the induction hypothesis), it should be that the clinched object is the unique available object in $O_{i_t}(\theta_{i_t})$ (by definition of clinching). Hence, i_t clinches the last remaining available object at $O_{i_t}(\theta_{i_t})$ at step t, which in turn implies that at step $t+1$, no object in $O_{i_t}(\theta_{i_t})$ is available. Since, $\bar{O}_t \subsetneq \bar{O}_{t+1} = O_{i_{t+1}}(\theta_{i_{t+1}})$, we conclude that $O_{i_{t+1}}(\theta_{i_{t+1}}) \setminus \bar{O}_t$ is a singleton. In other words, there is a unique available object in $O_{i_{t+1}}(\theta_{i_{t+1}})$, which should be clinched by i_{t+1} at step $t+1$. This completes the proof of the 'if' part.

We now prove the 'only if' part. The proof is by contraposition. Suppose, there is an $n \in \{1, 2, \ldots, N\}$, such that $\bar{O}_n \setminus \bar{O}_{n+1} \neq \emptyset$. Let us pick the smallest such n.

With the similar induction proof as in the 'if' part, but only applied up until to step $t = n$, we can see that at step $t+1$ all objects $\bar{O}_t$ are unavailable, and all objects in $\mathcal{O} \setminus \bar{O}_t$ are available. Note that because i_{t+1} is the dictator at step $t+1$, no individual other than i_{t+1} can clinch an object among $\mathcal{O} \setminus \bar{O}_t$ at this step. Thus, i_{t+1} is the only potential candidate for being a clincher. Yet, we will show that i_{t+1} too cannot clinch an object among $\mathcal{O} \setminus \bar{O}_t$. By the choice of t, we know that $\bar{O}_t \setminus \bar{O}_{t+1} \neq \emptyset$. Since $\bar{O}_t \subsetneq \bar{O}_{t+1} = O_{i_{t+1}}(\theta_{i_{t+1}})$, there should be at least two available objects in $O_{i_{t+1}}(\theta_{i_{t+1}}) \setminus \bar{O}_{t+1}$. Hence, i_{t+1} does not clinch any object among $\mathcal{O} \setminus \bar{O}_t$. This completes the proof of the 'only if' part of the Lemma. $\square$

Computing the proportion of problems that satisfy the conditions of Lemma 3 is a tractable combinatorial exercise. This proportion goes to zero as N becomes large, which proves Proposition 9.