Market Design Under Constraints

By

Joseph Root

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Economics

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Associate Professor Haluk Ergin, Chair
Professor David S. Ahn
Professor Chris Shannon

Summer 2020

Market Design Under Constraints

Abstract

Market Design Under Constraints

By

Joseph Root

Doctor of Philosophy in Economics

University of California, Berkeley

Associate Professor Haluk Ergin, Chair

With the development of market design and the increasing number of applications from school choice to kidney exchange has come the need for flexibility. The wide variety of practical settings calls for a set of tools which can be used broadly to incorporate market specific details. This dissertation collects three efforts to this end. The first two chapters concern the development of incentive compatible and Pareto efficient mechanisms in a setting where constraints are taken very broadly. The third abandons Pareto efficiency in favor of stability.

In the first chapter, coauthored with David S. Ahn, we study private-good allocation mechanisms where an arbitrary constraint delimits the set of feasible joint allocations. This generality provides a unified perspective over several prominent examples that can be parameterized as constraints in this model, including house allocation, roommate assignment, and social choice. We first characterize the set of two-agent strategy-proof and Pareto efficient mechanisms, showing that every mechanism is a "local dictatorship." For more than two agents, we leverage this result to provide a new characterization of group strategy-proofness. In particular, an N-agent mechanism is group strategy-proof if and only if all its two-agent marginal mechanisms (defined by holding fixed all but two agents' preferences) are individually strategy-proof and Pareto efficient. To illustrate their usefulness, we apply these results to the roommates problem to discover the novel finding that all group strategy-proof and Pareto efficient mechanisms are generalized serial dictatorships, a new class of mechanisms. Our results also yield a simple new proof of the Gibbard–Satterthwaite Theorem.

The second chapter, coauthored with David S. Ahn, takes a more concrete approach. In the same setting as chapter 1, we introduce a large subclass of mechanisms which we dub "constraint-traversing" and explore their properties. In particular, we provide two weak conditions – forward consistency and backward consistency – which, if satisfied, guarantee that a mechanism is group strategy-proof and Pareto efficient.

We illustrate the usefulness of this approach by deriving the set of 3-agent and 3-object house allocation mechanisms (already characterized by Pycia and Ünver (2017)). In addition, we demonstrate that these conditions can be equally applied to many "nearby" problems that would otherwise be intractable. Constraint-traversing mechanisms have a number of convenient properties. First, group strategy-proofness implies Pareto efficiency. Second, the marginal mechanisms of any constraint-traversing mechanism is also constraint-traversing.

In the final chapter, I consider stability in a two-sided matching context rather than incentive compatibility and Pareto efficiency in allocation mechanisms. I introduce a unified framework for studying two-sided matching problems with constraints. I introduce a matching algorithm called the constrained cumulative deferred acceptance algorithm capable of accommodating a wide variety of constraints. Like the deferred acceptance algorithm, one side of the market makes proposals to another. A "constraint correspondence" dynamically limits the choices of the receiving side in order to enforce that the ultimate match satisfies the constraint. If the constraint correspondence satisfies a "generalized substitutes" condition, the ultimate match will be constrained stable in the sense that satisfying any blocking pair would lead to a violation of the constraint. I provide two further conditions, "aggregate monotonicity" and "constraint IIA," on the constraint correspondence which ensure the constrained cumulative deferred acceptance algorithm implements a strategy-proof mechanism. Finally, I study the comparative statics of constraint correspondences.

Acknowledgements

I thank my dissertation committee Haluk Ergin, David Ahn and Chris Shannon for their guidance throughout my time at Berkeley. In particular, I would like to acknowledge Haluk Ergin for his integrity, generosity and for introducing me to Market Design. Our long afternoon meetings taught me to find the simplest version of every problem. I thank David Ahn for his encouragement, dedication and clarity of thought. Our work together made my time in graduate school deeply fulfilling. I would also like to thank Chris Shannon for her proactive and persistent encouragement.

I am deeply grateful to my wife Allyson for being my partner in curiosity and for her insight and interest during our long walks in Berkeley. I would like to acknowledege the support of my friends and family. My Dad, Donald Root, who never tired of hearing about my latest project and helped renew my curiosity. My mom, Jennifer Kovarik for her love and support. My brother Graham Root and my friends Collin Grenfell and Matt Kaye for finding a way to visit every time I needed a break.

The friendship of Brent Delbridge and my classmates Maxim Massenkoff, Caroline Le Pennec-Caldichoury, Leah Shiferaw and Katerina Jensen made my years in Berkeley wonderful.

I would also like to thank the people who helped shape my early career and decision to pursue research in economics. Stefano Dellavigna and Barry Eichengreen both generously guided my introduction to economics with patience. Richard Hornbeck and Dave Donaldson hired me as one of their first research assistants and gave me the opportunity to provide input and analysis in areas ranging from the effect of the Ogallala Aquifer to the impact of the great Boston fire of 1872 to the effect of the railroads on American economic growth. While I eventually returned to theory, I am a better economist for my experience in applied work with these outstanding people. Also important to my early years as an economist were Ray Kleunder and Adrienne Sabety. We began our lives as serious independent thinkers in our many discussions over coffee, dinner and beer in Cambridge.

Contents

Chapter 1

Preface

In this chapter, David S. Ahn and I explore constrained allocation under constraints. We aim to characterize the set of mechanisms which satisfy incentive and efficiency conditions. Specifically, we focus on mechanisms that are group strategy-proof and Pareto efficient. We arrive at a complete characterization for 2 agents and an indirect characterization when there are more than two agents. Applied to a variety of special constraints, this characterization yields new results. In the next chapter we use this characterization to introduce a large class of mechanisms motivated by our findings.

Incentives and Efficiency in Constrained Allocation Mechanisms

Joseph Root[1] and David S. Ahn[2]

1.1 Introduction

Many market design problems involve constraints. School choice assignments must ensure quotas of low-income students are satisfied at high-performing schools. Medical residency assignments must place enough doctors in rural areas. The allocation of radio frequency in spectrum auctions must satisfy a large number of complicated engineering conditions to ensure minimal cross-channel interference.

Although successful ad hoc approaches have been tailored for particular problems, to date there is little general understanding of how constraints affect efficiency and incentives, the two classic criteria for implementation. Theoretically, a unified approach would enable analytical insights to be shared between contexts. Practically, a flexible theory of constraints for market design would greatly expand applicability. Real-world problems involve many considerations that are difficult to anticipate. The tools of market design should be general enough to accommodate these considerations.

We develop a model of object allocation with private values for completely general constraints. A finite number of objects are allocated to a finite number of agents and an arbitrary constraint circumscribes the set of feasible social allocations. Each agent has strict preferences over the objects assigned to her, but is indifferent to others' assignments.

While other agents' assignments have no direct effect on one's well-being, those assignments do limit the profiles of allocations that are jointly feasible. Obviously, the assignment of a house to another agent precludes my consumption of that house.

[1]Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: jroot@econ.berkeley.edu

[2]Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: jroot@econ.berkeley.edu

So even with purely private values, constraints introduce linkage across agents' allocations. Each agent i is indirectly concerned with any other j's assignment, not because i cares about j's consumption, but rather because j's assignment will limit the set of objects for i that are jointly feasible with the j's assignment. Our goal is to study the set of incentive compatible and efficient mechanisms for a fixed arbitrary constraint. In addition, we aim to study how different features of a constraint make it amenable for implementation, that is, to understand what kinds of constraints yield what kinds of truthful and efficient mechanisms. For any constraint on the set of feasible allocations, our main findings characterize the entire class of mechanisms that are immune to manipulation by any group of agents yet still yield Pareto efficient outcomes.

Beyond its practical benefits, a general theory of constrained allocation yields some surprising theoretical insights. Several prominent problems which at first glance may appear unconstrained and unrelated can be neatly expressed as special constraints of our model. For example, the classical social choice problem corresponds to the special constraint of our model where all agents are constrained to consume the same object.[3] From this perspective, the social choice problem presents itself as a special constrained private-goods allocation problem. In fact, a corollary application of our results is the Gibbard–Satterthwaite Theorem: that all strategy-proof social choice mechanisms are dictatorial. With this novel presentation of social choice as a constraint, we can now sensibly formulate and prove a converse to Gibbard–Satterthwaite: under what conditions does the constraint admit any non-dictatorial mechanism?

Another prominent application of our theory is to house allocation, where a finite number of indivisible objects must be assigned to agents with unit-demand. Expressed this way, the house allocation problem is almost the opposite of the social choice problem: no two agents can be assigned the same object. Recently, Pycia and Ünver (2017) provided a full characterization of the group strategy-proof[4] and Pareto efficient house allocation mechanisms, building on earlier work by Papái (2000). In an earlier version of this paper, we show how to use our results to replicate Pycia and Ünver (2017) for a small number of agents.[5]

A third prominent problem that can be expressed as a constraint is the roommates problem, where an even number of agents need to match into pairs. In this case, the "objects" are the other agents and the constraint requires that: first, no agent is matched to herself; and second, if i is assigned to j then j is commensurately assigned to i. In contrast to the previous two applications, to our knowledge no general characterization of the incentive compatible, efficient mechanisms had yet been discovered. As an application of our results, we provide such a characterization. We show that

[3]The term "object" is figurative. In social choice, the objects are usually policy choices or political candidates.

[4]Roughly, a mechanism is group strategy-proof if no coalition of agents can jointly misreport their preferences, without harming anyone in the group and making at least one agent strictly better off.

[5]The argument constructs a tedious change of variables to parameterize the Pycia and Ünver (2017) as a special case of our general formulae in the three-agent case. Details are available from the authors on request.

all group strategy-proof and Pareto efficient roommates mechanisms are "generalized serial dictatorships," a class of mechanisms we will formally introduce later.[6] The fact that our results are useful in understanding and proving results across some well-known problems is a fortunate side-effect of the model's generality.

These examples illustrate a key conceptual contribution of our paper: to provide a novel framework to unify positive and negative results across these applications, tying together seemingly disparate environments and results by viewing them as different constraints on the image rather than through restrictions of preferences on the domain. Traditionally, positive results in specific environments are seen as escaping the impossibility of the Gibbard–Sattertwhaite Theorem by restricting preferences in the *domain* of the mechanism to convenient special cases, such as assuming single-peaked rankings or quasi-linear preferences. In our model, we can provide a different reconciliation of these positive results by interpreting these environments as relaxing constraints in the *image* of the mechanism: outside of the Arrovian social choice problem, all agents need not consume the same object and instead there is room for compromise to yield mechanisms beyond dictatorship. The "diagonal" constraint implicit in the social choice problem generates maximal tension between efficiency and incentives, while other constraints allow more scope for their coexistence. Our model explicitly exposes this tension, and our results characterize the scope for positive incentive-compatible implementation of efficient outcomes when this tension is relaxed. This provides a deeper understanding of why certain environments like social choice admit so few good mechanisms while other environments like house allocation admit a broad variety of good mechanisms.

Despite allowing for complete generality in the constraint, we fully characterization all mechanisms that satisfy standard incentive and efficiency desiderata. We start by considering two-agent environments. This case admits a surprisingly parsimonious characterization of the set of individually strategy-proof and Pareto efficient mechanisms for all constraints. We show that all individually strategy-proof and Pareto efficient mechanisms are "local dictatorships" in which the set of infeasible allocations is partitioned into two regions and each region is assigned a local dictator. For a given preference profile, the agents' top choices determine some (possibly infeasible) social allocation. If this allocation is feasible, the mechanism assigns it. Otherwise, it is infeasible and there is a local dictator assigned to the allocation. The non-dictator is assigned their favorite object compatible with the dictator's top object. However, not all partitions will maintain efficiency and incentive compatibility. Instead, some structure is required of the partition to ensure these desiderata are maintained. We show that every constraint can have its infeasible allocations "block diagonalized" to yield an immediate characterization of the partitions that do yield desirable mechanisms. Every block must be assigned to a single agent as the local dictator. So the number of strategy-proof and Pareto efficient mechanisms is determined entirely by the number

[6]In common with standard serial dictatorship, there is a sequence of dictators and each dictator picks her favorite object among those that are possibly feasible with the choices of earlier dictators. In contrast to standard serial dictatorship, our generalized version allows the order of subsequent dictators to depend on the choices of earlier dictators, rather than being locked in a fixed order.

of blocks allowed by the constraint.

With three or more agents, the set of individually strategy-proof and Pareto efficient mechanisms no longer admits such a straightforward characterization. Indeed, even for the classic house allocation setting, the collection of all such mechanisms is still unknown. Nevertheless, if we strengthen our incentive compatibility condition to group strategy-proofness, we can leverage the two-agent results to get a novel recursive characterization for the multi-agent case. Group strategy-proofness requires that no group of agents can ever collectively misreport their preferences so that all agents in the group are weakly better off and at least one agent is strictly better off. Our central observation is that group strategy-proof mechanisms have the convenient property that we can restrict attention to a subset of agents, fixing a preference profile of everyone else, to get a new group strategy-proof mechanism for the subset. We call these the "marginal mechanisms." Importantly, the properties of just the two agent marginal mechanisms are enough to capture the group incentives of the entire mechanism: if all two-agent marginal mechanisms are Pareto efficient and individually strategy-proof, then the full mechanism is group strategy-proof. This discovery is especially useful given our explicit characterization of two-agent mechanisms. The two-agent mechanisms of our first result are therefore the "building blocks" of all group strategy-proof mechanisms with many agents.

Beyond its analytical power, group strategy-proofness is substantively natural for a number of reasons. First, we show that, for any constraint, group strategy-proofness is equivalent to individual strategy-proofness and a classic normative condition called "nonbossiness"[7]. In bossy mechanisms, agents can manipulate the outcome of other agents without affecting their own allocation. Therefore, the marginal power of restricting attention to group strategy-proofness, relative to requiring only individual strategy-proofness, is simply to rule out such bossy mechanisms. So the gap between group and individual incentives boils down to whether one agent is allowed to alter another's outcome while not changing her own outcome. Second, in practice, incentive problems have been highly detrimental to the practical appeal of mechanisms. Violations in strategy-proofness of the Boston mechanism lead to severe inequality between "sophisticated" agents who knew how to game the system and "naive" agents who didn't. Ultimately, the mechanism was replaced in favor a strategy-proof mechanism (Abdulkadiroğlu, Pathak, Roth, and Sonmez 2006). The Vickrey-Clarke-Groves mechanism, despite its attractive individual incentives, has largely not been implemented in practice, in part because of its susceptibility to group manipulation (Rothkopf 2007). We therefore believe that mechanisms with strong group incentives are especially useful for practical considerations. In addition, group strategy-proofness is among the most demanding incentive conditions in the literature, and this benchmark should be established to understand the gains to efficiency from demanding weaker incentive conditions like Bayesian implementation. Finally, group strategy-proofness has been long studied in other environments, and especially for the house allocation problem, so using this as

[7]To our knowledge, nonbossiness was first introduced by (Satterthwaite and Sonnenschein 1981).

our incentive condition facilitates comparisons with earlier results. That all said, our focus on Pareto-efficiency and group strategy-proofness rules out some practical mechanisms. Deferred acceptance, for example, is not Pareto efficient, is not individually strategy-proof for the accepting side, and is not group strategy-proof for the proposing side.

1.1.1 Literature Review

To our knowledge, this paper is the first to identify the entire set of mechanisms that satisfy criteria regarding incentives and efficiency for an arbitrary constraint in our general allocation problem. However, several papers study mechanisms for specific constraints in particular environments. One such environment is the two-sided matching problem with distributional constraints, where for example there is a cap on the number of medical residents assigned to hospitals in a certain area. The two-sided matching problem can be expressed as a constraint in our more general model, and distributional constraints can be expressed as a further sharpening of that constraint.[8] A series of papers summarized Kamada and Kojima (2017a) study the two-sided matching problem with distributional constraints, with a primary focus on understanding stability.[9] In the two-sided matching problem, stability is the primary normative concern since the ubiquitous deferred-acceptance mechanism is known to be neither strategy-proof nor Pareto efficient. While specific mechanisms are shown to work well for specific classes of constraints, a general accounting for the class of all mechanisms is still outstanding. In principle, our results applied to this problem would characterize the set of all group strategy-proof and Pareto efficient mechanisms. That said, our results are exclusively about incentives and efficiency, and we have little to directly say about stability. This is partly because, as a concept, stability is only sensible and well-defined in particular examples of our environment such as two-sided matching.

Another example of a particular environment with a constraint on allocations is the house allocation problem, although it is not often thought of as a constrained problem. Abdulkadiroğlu and Sönmez (1999) and Papái (2000) construct classes of group strategy-proof and Pareto efficient mechanisms that are strictly larger than two classic examples of group strategy-proof and Pareto efficient mechanisms for house allocation: top trading cycles, attributed to David Gale by Shapley and Scarf (1974) and shown to have these desirable features by Bird (1984), and serial dictatorship, analyzed comprehensively by Svensson (1994) and Svensson (1999), which obviously has these features. A general characterization had remained a long-standing problem

[8] More precisely, the two-sided matching problem can be modeled by making the set of objects equal to the union of agents from both sides of the market with the constraint that each agent is assigned to an agent in the opposite side and that, if agent i is matched to agent j then j should also be matched to i.

[9] Work in this literature includes contributions by Hafalir, Yenmez, and Yildirim (2013), by Ehlers, Hafalir, Yenmez, and Yildirim (2013), by Kamada and Kojima (2015), by Kamada and Kojima (2017b), and by Kamada and Kojima (2018).

until Pycia and Ünver (2017) recently provided an impressive full description of all group strategy-proof and Pareto efficient mechanisms mechanisms. These are exactly the normative criteria explored in this paper, and in fact Pycia and Ünver (2017) helped inspire this paper by demonstrating a general characterization of these criteria is even attainable for an important problem like house allocation. House allocation problems are a special constraint in our model, where $a_i \neq a_j$ is required whenever $i \neq j$. That is, our characterization when applied to this constraint also provides another parameterization of mechanisms in Pycia and Unver. We explicitly verify the connection between the two characterizations in the three-house case, and believe the general change of variables between the two formulations is feasible but would be very tedious.

While incentives and efficiency are relatively well-understood for two-sided matching and for house allocation, one-sided matching such as in the classic problem of pairing roommates into dormitory rooms has demonstrated itself to be much more intractable. This is in large part because one-sided environments may fail to yield a stable match, as originally observed by Gale and Shapley (1962) in the same article introducing their eponymous algorithm for stable two-sided matching. Since then, a very large literature in operations research and computer science, starting with Irving (1985), tries to find efficient algorithms to find stable matchings when they exist. This specific computational problem has become so well-studied that it is now called the "stable roommates problem." In contrast, there seems to be almost no discussion of incentives and efficiency for the roommates problem.[10] An application of our main results yields a characterization of group strategy-proof and Pareto efficient mechanisms for the roommates problem, which turn out to be the family of generalized serial dictatorships that we introduce in this paper. To our knowledge, this is a new observation and, analogous to the characterization theorem by Pycia and Ünver (2017) for house allocation or to the Gibbard–Satterthwaite Theorem for social choice, establishes the characterization of group strategy-proofness and Pareto efficiency for the roommates problem.

A final notable special constraint in our environment is the classic Arrovian social choice model. The first result studying incentives and efficiency was the celebrated negative finding by Gibbard (1973) and Satterthwaite (1975), which initiated the field of implementation theory. Here, the classic Arrovian social choice environment in which the Gibbard–Satterthwaite Theorem is cast corresponds to the case where all agents must be assigned the same common outcome. That is, social choice corresponds to the constraint that $a_i = a_j$ for all agents i, j. Viewed in this way, the social choice constraint is almost the opposite of the house allocation constraint. We derive the Gibbard-Sattherhwaite Theorem as a corollary of our main characterization. This provides a novel perspective on the classic result by casting light on the implications of constraining allocations so that all agents consume a common object. Our perspective allows us to understand the Gibbard–Satterthwaite Theorem as a consequence of the restrictiveness

[10]The one exception we found was a working paper by Abraham and Manlove (2004) that studies the computational hardness of finding Pareto optimal matches for the roommates problem.

of the constraint. Correspondingly, our perspective also offers a novel escape from the assumptions of the Gibbard–Satterthwaite Theorem, namely relaxing the social choice constraint. This escape is meaningful only when Arrovian social choice is framed as a special case of private good economies. In fact, this framing allows us to generalize the Gibbard–Satterthwaite Theorem in our environment: we completely characterize the constraints where only serial dictatorships are group strategy-proof, finding the social choice constraint as a particular example. It is interesting that social choice can be cast as a special case of our model with the particular diagonal restriction on allocations, since private-goods economies are usually viewed as a special case of social choice with a particular restriction on preferences.

Our general environment with private goods was also recently studied by Barberà, Berga, and Moreno (2016) from a social choice perspective. They work focuses on the richness of preferences for a social choice function, that is, it focuses on the richness of the *domain* of preference. Throughout our paper, by contrast, we allow no restrictions on preferences and assume that mechanisms will find allocations for all preference profiles. Instead of considering restrictions on the domain, we complement Barberà, Berga, and Moreno (2016) by considering different constraints on the *image* of allocations that are feasible for a mechanism.

Our different focus on constraints on allocations, rather than on restrictions over preferences, stems partly from our different objectives. Barberà, Berga, and Moreno (2016) are primarily concerned with the relationship between group and individual incentives. Their main result reveals an important connection between group and individual strategy-proofness when the space of admissible preferences is sufficiently rich.[11] In contrast, our aim is not to relate different axioms for strategy-proofness, but rather to characterize the entire space of mechanisms that satisfy the fixed axiom of group strategy-proofness. Our main results examine the structure of the constraint to describe the structure of the group strategy-proof mechanisms. That is, our objective is not to relate strategy-proofness to other normative conditions like nonbossiness or monotonicity, but rather to relate the structure of group strategy-proof mechanisms to the structure of the constraint. Our results address concerns like how the space of strategy-proof mechanisms changes when constraints are relaxed or tightened. Of course, an improved understanding of how group strategy-proofness relates to other natural conditions can only be helpful. In fact, a key lemma in proving our characterization is to observe a tight relationship between group strategy-proofness, individual strategy-proofness and nonbossiness, and Maskin monotonicity. So our development owes a debt to these earlier realizations. However, our lemma is still distinct from these earlier observations in both substance and message, as we will explain after formally introducing the result.

Finally, a more distant body of work on random allocation tests whether a random allocation is a convex combination of deterministic allocations satisfying a fixed

[11] This complements a similar connection between group and invidiual incentives for classic Arrovian environments, discovered by the same authors (Barberà, Berga, and Moreno 2010) and by Le Breton and Zaporozhets (2009).

constraint (Balbuzanov 2019, Budish, Che, Kojima, and Milgrom 2013), extending the fairness gains of the random assignment mechanisms introduced by Bogomolnaia and Moulin (1990) to constrained environments. We focus on deterministic mechanisms, so as far as we can see our results have no direct relationship to this literature.

1.2 Model

We begin by introducing primitives. Let N be a finite set of **agents** and $\mathcal{O}$ be a finite set of **objects**. We use the term "object" because of our leading examples, but note that these are not necessarily physical objects, but can be political candidates, roommates, and so on. Define $\mathcal{A} = \mathcal{O}^N$ to be the set of all possible allocations of objects to agents. Equivalently, $\mathcal{A}$ is also the set of maps $\mu : N \to \mathcal{O}$ and we switch to this perspective when it is more useful. A **suballocation** is a map $\sigma : M \to \mathcal{O}$ where $M \subset N$. Let $\mathcal{S}$ denote the set of suballocations. Our task is to assign objects to agents in a way that is consistent with an exogenous *constraint* which reflects the set of feasible allocations for a particular application. Importantly, the constraint is exogenous to the problem. It is given to the mechanism designer as a fixed set of feasible outcomes. Formally, we are given a nonempty **constraint** $C \subset \mathcal{A}$ and $(a_i)_{i \in N} \in C$ means that it is feasible to allocate each agent i the object a_i simultaneously. Notice that since we place no restrictions on the constraint, it is without loss of generality to have a common set of objects for all agents because if each agent has her own set of objects then one could add the constraint that all feasible allocations cannot assign these objects to other agents.[12] Agents have strict preferences over the *objects* and are assumed to be indifferent between any two allocations in which they receive the same object. We will use P to denote the set of strict preferences (i.e. linear orders) on $\mathcal{O}$ and $\mathscr{P} = P^N$ to denote the set of preference profiles.[13] Our primary object of interest in this paper is a **feasible mechanism**, which is simply a map $f : \mathscr{P} \to C$. Our task will be to find feasible mechanisms satisfying desirable conditions regarding incentives and efficiency, to be formally introduced in the sequel.

Some well-known problems can be expressed as special constraints in this model:

- House Allocation: A finite number of houses must be distributed to a finite number of agents. The houses cannot be shared so no two agents can be allocated the same one. This gives rise to the constraint

$$C = \{(a_i)_{i \in N} \,|\, a_i \neq a_j \text{ when } i \neq j\}.$$

  This setting has been the subject of considerable interest since at least Shapley and Scarf (1974). Two prominent mechanisms used in practice are Gale's top trading cycles algorithm and Gale and Shapley's deferred acceptance algorithm (with priorities for houses).

[12] More precisely, let $\mathcal{O} = \sqcup \mathcal{O}_i$ and define C_{new} by $(a_i)_{i \in N} \in C_{new}$ if and only if $(a_i)_{i \in N} \in C$

[13] A binary relation $B \subset \mathcal{O} \times \mathcal{O}$ is a linear order if it is complete, transitive, and antisymmetric

- Roommates Problem: Universities are often tasked with assigning students into shared dormitory rooms. Assuming N is even, this problem can be captured in our environment by setting $\mathcal{O} = N$ and imposing the constraint

$$C = \{\mu : N \to N \,|\, \mu^2 = id \text{ and } \mu(i) \neq i \text{ for all } i\}.$$

  The first condition requires that if i is assigned roommate j then j is also assigned i and the second condition requires that all agents are assigned a roommate.

- Social Choice: If the constraint specifies that all agents receive the same object (without specifying ex-ante which object will be chosen) we get the classical version of the social choice problem[14]. Specifically, if

$$C = \{(a_i)_{i \in N} \,|\, a_i = a_j \text{ for all } i, j\}$$

  the constraint requires that all agents be given the same social choice, but which outcome is chosen is a function of the mechanism.

Our model is able to accommodate these examples as special cases because of its generality in admitting arbitrary constraints. We will have more explicit analyses of these examples later in the paper.

Before moving on, we record here some notation used throughout the paper. For any subset $M \subset N$, given a preference profile $\succsim = (\succsim_i)_{i \in N} \in \mathscr{P}$ and a profile of alternative preferences for agents in M, $(\succsim'_j)_{j \in M}$, we will write $(\succsim'_M, \succsim_{-M})$ to refer to the profile in which an agent j from M reports $\succsim'_j$ and any agent i from M^c reports $\succsim_i$. We will often want to consider how a mechanism f changes when a few agents change their preferences, that is the difference between $f(\succsim)$ and $f(\succsim'_M, \succsim_{-M})$. When the initial preference profile $\succsim$ is clear, we will sometimes write $\succsim_-$ instead of $\succsim_{-M}$. Given a constraint $C \subset \mathcal{A}$ and a subset of agents $M \subset N$, let $C^M = \{\mu : M \to \mathcal{O} \,|\, \exists b \in C \text{ s.t. } b_i = \mu(i)\, \forall i \in M\}$ which we will call the **projection of** C **on** M. An element of C^M will be referred to as a **feasible suballocation** for agents in M. If $\mu : M \to \mathcal{O}$ and $\mu' : M' \to \mathcal{O}$ are suballocations with $M \subset M'$ which agree on their shared domain, μ' is called a **extension** of μ. If μ' is a feasible suballocation (which of course implies that μ is) then μ' is called a **feasible extension** of μ. If μ' assigns an object to each agent, it is called a **complete extension** of μ. Given a feasible suballocation μ, we will let $C(\mu)$ denote the set of complete and feasible extensions of μ. For any agent i, let $\pi_i : \mathcal{A} \to \mathcal{O}$ be the projection map so that given an allocation $(a_j)_{j \in N}$, $\pi_i a = a_i$ and for a set of allocations $B \subset \mathcal{A}$, we have $\pi_i B = \{a \in \mathcal{O} \,|\, \text{ there is a } b \in B \text{ with } \pi_i b = a\}$. For $x \in \mathcal{O}$ and $\succsim_i \in P$, define $LC_{\succsim_i}(x) = \{y \in \mathcal{O} \,|\, y \prec_i x\}$ be the (strict) **lower contour set** of x at $\succsim_i$. Likewise, $UC_{\succsim_i}(x) = \{y \in \mathcal{O} \,|\, y \succ_i x\}$ is the (strict) **upper contour set** of x at $\succsim_i$. For a preference $\succsim_i$, define $\tau_n(\succsim_i)$ as the nth top choice under $\succsim_i$. Likewise, for any preference profile $\succsim$, define $\tau_n(\succsim)$ as the allocation in which each agent gets their nth top choice. To save on notation, we will often omit the subscript when referring to

[14] See Barberà (2001) for a general statement of the social choice problem with restricted domains.

the top choice (i.e. writing $\tau(\succsim)$ to mean $\tau_1(\succsim)$). We will use $\bar{C}$ to denote the set of infeasible allocations.

In practice, mechanisms are often designed to satisfy efficiency and incentive properties. Here are several well-known desiderata for allocation mechanisms.

Definition 1. A mechanism $f : \mathscr{P} \to C$ is

1. **strategy-proof** if, for every $i \in N$ and every $\succsim \in \mathscr{P}$,

$$f_i(\succsim) \succsim_i f_i(\succsim'_i, \succsim_{-i})$$

for all $\succsim'_i \in P$. That is, truth-telling is a weakly dominant strategy.

2. **group strategy-proof** if, for every $\succsim \in \mathscr{P}$ and every $M \subset N$, there is no $\succsim'_M$ such that

   (a) $f_j(\succsim'_M, \succsim_{-M}) \succsim_j f_j(\succsim)$ for all $j \in M$;
   (b) $f_k(\succsim'_M, \succsim_{-M}) \succ_k f_k(\succsim)$ for at least one $k \in M$.

3. **weakly group strategy-proof** if, for every $\succsim \in \mathscr{P}$ and every $M \subset N$, there is no $\succsim'_M$ such that

$$f_j(\succsim'_M, \succsim_{-M}) \succ_j f_j(\succsim) \text{ for all } j \in M.$$

4. **Pareto efficient** if there is no allocation $a \in C$ and preference profile $\succsim$ such that $a \neq f(\succsim)$ and $a_j \succsim_j f(\succsim)$ for all j.

5. **nonbossy** if, for all $\succsim \in \mathscr{P}$,

$$f_i(\succsim'_i, \succsim_{-i}) = f_i(\succsim) \implies f(\succsim'_i, \succsim_{-i}) = f(\succsim).$$

6. **Maskin monotonic** if, for all $\succsim, \succsim' \in \mathscr{P}$,

$$LC_{\succsim'_i}[f_i(\succsim)] \supset LC_{\succsim_i}[f_i(\succsim)] \text{ for all } i \implies f(\succsim') = f(\succsim).$$

Strategy-proofness requires that for every agent i and every possible profile of preferences for the other agents, i cannot improve her outcome by misreporting her preference. Group strategy-proofness is similar except that it requires that no group can collectively misreport their preferences without hurting anyone while strictly benefiting at least one agent. This is often called "strong group strategy-proofness" to contrast it with weak group strategy-proofness which requires that any deviating coalition make all its agents strictly better off. Pareto efficiency might also be called "constrained efficiency" since it requires that for every preference profile f selects a feasible allocation such that no other feasible allocation can improve (at least weakly) all agents outcomes. Pareto efficiency is also sometimes called "unanamity" in the literature. Nonbossiness simply requires that no agent can exert influence on another agent without affecting her own outcome. Finally, Maskin monotonicity is the seemingly weak condition that whenever

an allocation is chosen at a given preference profile, if all agents instead report a different profile in which their respective allocations have improved relative to all other allocations, then f should maintain the same outcome. This condition was famously shown to be necessary for Nash implementation by Maskin (1999).

A useful observation in building our results is the following equivalence across these conditions. We present this lemma explicitly because it is of some independent interest and to explain how this part of our argument relates to earlier observations.

Proposition 1. *If $f : \mathscr{P} \to \mathcal{A}$ the following are equivalent:*

1. *f is group strategy-proof.*

2. *f is strategy-proof and nonbossy.*

3. *f is Maskin monotonic.*

The connection between individual and *weak* group-strategy proofness was examined in social choice environments by Le Breton and Zaporozhets (2009) and by Barberà, Berga, and Moreno (2010) and in private-goods environments such as ours by Barberà, Berga, and Moreno (2016), who prove that, when the domain of preference is sufficiently rich, weak group strategy-proofness is equivalent to individual strategy-proofness for a broad class of social choice functions satisfying generalizations of nonbossiness and Maskin monotonicity. An immediate difference is our use of strong rather than weak group strategy-proofness, which follows the literature on house allocation that also studies strong group strategy-proofness.[15] While perhaps a seemingly technical distinction, this is quite a substantively important departure from the weak concept. For example, deferred acceptance is only weakly group-strategyproof on the proposing side, but is not group strategy-proof in our stronger sense. Even ignoring the difference between weak and strong incentives, the theorem of Barberà, Berga, and Moreno (2016) bears no obvious relation to Proposition 1. The two results have very different aims and messages. Barberà, Berga, and Moreno (2016) take generalizations of Maskin monotonicity (that they call "joint monotonicity") and nonbossiness (that they call "respectfulness") as *assumptions* in their results and ask how large the domain of preferences must be to ensure group and individual incentives align. Our result generates nonbossiness and Maskin monotonicity as *implications* of group strategy-proofness for full preference domains, which is important in subsequent applications where we verify that a mechanism is group strategy-proof by testing that it is Maskin monotonic. On the other hand, we assume the domain of all strict preferences throughout this paper, and have nothing to say here about the consequences of restrictions on preferences.

The relationship between group strategy-proofness and Maskin monotonicity was first revealed by the proof of the Muller–Satterwthwaite Theorem, which proceeds by showing that either group or individual strategy-proofness is equivalent to Maskin

[15]For the specific problem of house allocation, the equivalence between (1) and (2) was first observed by Papái (2000).

monotonicity for the social choice problem (Muller and Satterthwaite 1977).[16] This equivalence between group strategy-proofness and Maskin monotonicity was then further demonstrated to hold for other problems as well, including for house allocation by Svensson (1999) and for two-sided matching by Takamiya (2001). Takamiya (2003) unified these observations in a general statement for all indivisible-good economies without externalities that also applies to our model, and should be credited for the equivalence between (1) and (3) in Proposition 1.

Group strategy-proofness requires that no group of agents can collectively misreport their preferences and benefit at least one agent without making anyone in the group worse off. One possible coalition is the grand coalition. Thus if f is group strategy-proof and $f(\succsim) = a$ for some profile $\succsim$, then a can never Pareto dominate $f(\succsim')$ for any other profile $\succsim'$, since all agents would collectively report $\succsim$.

Lemma 1. *If $f : \mathscr{P} \to \mathcal{A}$ is group strategy-proof then it is Pareto efficient on its image.*[17]

Having established this, the goal of this paper is to understand the correspondence between the primitives (the set of agents, objects, and the constraint) and the set of group strategy-proof, Pareto efficient mechanisms. We will denote the set of feasible group strategy-proof mechanisms which map into C, $GS(C)$.

1.3 Characterization Results

We begin by considering the two-agent case where we find an explicit characterization of the set of strategy-proof and Pareto efficient mechanisms for an arbitrary constraint. Each mechanism with these properties turns out to be a "local dictatorship." We then turn to the n-agent case where we show that an n-agent mechanism is group strategy-proof if and only if each 2-agent marginal mechanism is group strategy-proof.

1.3.1 Two Agents

Given just two agents, we will show that for every constraint the set of strategy-proof and Pareto efficient mechanisms corresponds exactly to the set of "local dictatorships" in which the set of infeasible allocations $\bar{C}$ is partitioned into two disjoint subsets and each agent is assigned a set. After the agents announce their preferences, if the allocation in which both agents get their top choice is feasible, the mechanism must pic this allocation by Pareto efficiency. Otherwise, it is infeasible to give both agents their top choices and one agent must compromise and consume a less-favored object. The agent who does not have to compromise is the "local dictator" and gets her top choice,

[16] Recall the Muller–Satterwhaite Theorem: all Maskin monotonic and surjective social choice functions are dictatorial.

[17] That is, if the constraint C is exactly $im(f)$.

and the "local compromiser" receives her favorite object among those that are jointly feasible with the local dictator's top choice.

One possible complication with this procedure is that there may be no object for the local compromiser that is jointly feasible with the local dictator's top choice. For example, if the local dictator at (x, y) is agent 1, and $(x, y') \notin C$ for all objects $y' \in \mathcal{O}$, then there is no choice for agent 2 that will allow agent 1 to consume her favorite object x. On the other hand, since agent 1 can never feasibly be assigned object x, it would seem that her preference for x is immaterial to the social choice. This turns out to be true, and we can ignore objects that are never assigned to an agent without loss of generality. To make this precise, for any constraint $C \subset \mathcal{O}^2$ let $R_1 = \{x \in \mathcal{O} \,|\, (x, y) \notin C \text{ for all } y \in \mathcal{O}\}$ and $R_2 = \{y \in \mathcal{O} \,|\, (x, y) \notin C \text{ for all } x \in \mathcal{O}\}$. In words, R_i is the set of objects which are always infeasible for agent i because there is no object a_{-i} for the other agent that will make the joint allocation (a_i, a_{-i}) feasible. More generally, we can likewise define R_i for any number of agents as the set of objects which are always infeasible to agent i no matter what objects are assigned to everyone else. Since these objects are immaterial to the agents, it would seem natural and would certainly be convenient if the ranking of always infeasible objects should have no effect on the outcome of a mechanism. The following lemma says exactly that.

Lemma 2. *Let C be a constraint for n agents. If $f : \mathscr{P} \to C$ is group strategy-proof and Pareto efficient and if $\succsim$ and $\succsim'$ are preference profiles in which for every i the relative ordering of elements in $\mathcal{O} \setminus R_i$ is unchanged then $f(\succsim) = f(\succsim')$*

Let $\bar{C}^* = \{(x, y) \,|\, (x, y) \notin C \text{ and } x \notin R_1, y \notin R_2\}$. That is, $\bar{C}^*$ is the set of infeasible allocations in which both agents could get the associated object for some choice of the other agents' object. As mentioned, all Pareto efficient mechanisms will assign top choices to both agents when doing so is feasible. The main job of a mechanism is to adjudicate the outcome when one agent must give up on her top choice. It turns out that strategy-proofness will demand a local dictator is determined as a function of only the agents' top objects. We prove this claim by taking an approach to strategy-proofness originally developed by Barberà (1983). This approach begins with the simple but deep observation that strategy-proof social choice functions can always be written as if an "option set" is available to player i as a function of everyone else's ($j \neq i$) report, and then i's allocation maximizes agent i's reported preference over that option set. We explicitly restate Barberà's observation for our environment of private goods, because we feel it is not as generally well-known as it should be and to acknowledge the role it plays in our argument. Let $P^{N-1} = \times_{j \neq i} P$ denote the space of preference profiles for all players beside agent i.

Lemma 3 (Barberà (1983))**.** *A mechanism $f : \mathcal{P} \to C$ is strategy-proof if and only if there exist nonempty correspondences $g_i : P^{N-1} \rightrightarrows \mathcal{O}$ such that, for all agents i,*

$$f_i(\succsim) = \max_{\succsim_i} g_i(\succsim_{-i})$$

With some work, Barberà's Lemma can be used to show that all strategy-proof and Pareto efficient two-agent mechanisms assign a local dictator who gets her top choice, and the assignment of dictatorship can depend only on the top choice for each agent. So such mechanisms can be described by coloring the set $\bar{C}^*$ with one color for the top-choice pairs where agent 1 is the local dictator and the other color for the top-choice pairs where j is the local dictator.

However, not all such colorings will be strategy-proof. For example, if agent 1 is the local dictator when (a, b) are the top choices and agent 2 is the local dictator at (a, b'), then agent 2 may want to misreport her top choice as b' even in situations where b is actually her top choice because she gets dictatorship power by misreporting. The coloring of the infeasible set $\bar{C}^*$ will have to satisfy some restrictions, which motivates the following constructions. Define the binary relation B on $\bar{C}^*$ by $(a, b)B(a', b')$ if $a = a'$ or $b = b'$. Two allocations are related by B if (at least) one agent gets the same object in both allocations. Now if $(a, b)B(a', b')$, then the example above suggests that the same agents must be assigned as the dictator in both cases, to prevent the situation where one agent can move from being the local compromiser to being the local dictator by individually misreporting her top object. This relation must hold across pairs of top choices that are even indirectly linked, so common assignment of local dictatorship must also hold transitively across B. Let T be the transitive closure of B.[18] Since B is reflexive and symmetric, it can easily be shown that T is an equivalence relation.[19] As an equivalence relation on a finite set, it can be expressed as a partition with a finite number of equivalence classes $E_1, E_2, \ldots E_p$, where $(a, b)T(a', b')$ if and only if (a, b) and (a', b') are both in some E_i. We will refer to the equivalence classes of T as the **blocks** of $\bar{C}^*$.

Figure 1.1 illustrates an example of the relation T for a specific constraint. The top-left panel shows the constraint; grey cells are infeasible allocations. Panel (B) permutes $R_1 = \{a_4\}$ and $R_2 = \{a_4, a_6\}$ to the top and left most objects. In panel (C), a particular 4-element block of $\bar{C}^*$ consisting of (a_2, a_1), (a_2, a_3), (a_6, a_3), and (a_6, a_8) is shaded black. No element of the grey set is related by B to any member of $\bar{C}^*$ which is not also shaded black. Since the order of objects is not important, we can permute the rows and columns to display the equivalence classes more easily. Hence in panel (D), we again permute the objects. As we can now easily see there are three equivalences classes of T which are indicated as E_1, E_2 and E_3. We can then assign a dictator to each block independently as described below.

Let $C^1(b) = \{a \in \mathcal{O} \,|\, (a, b) \in C\}$ and likewise $C^2(a) = \{b \in \mathcal{O} \,|\, (a, b) \in C\}$. A mechanism $f : \mathscr{P}^2 \to C$ is called a **local dictatorship** if each block E_i of $\bar{C}^*$ is

[18] The transitive closure is the minimum binary relation containing B which is transitive.

[19] It is reflexive because B is. To see that it is symmetric, if we have $(a, b)T(a', b')$ since $\bar{C}^*$ is finite, there are $(a_1, b_1), \ldots (a_n, b_n)$ such that $(a, b)B(a_1, b_1)B \ldots B(a_n, b_n)B(a', b')$. By reversing all these, we see that $(a', b')T(a, b)$.

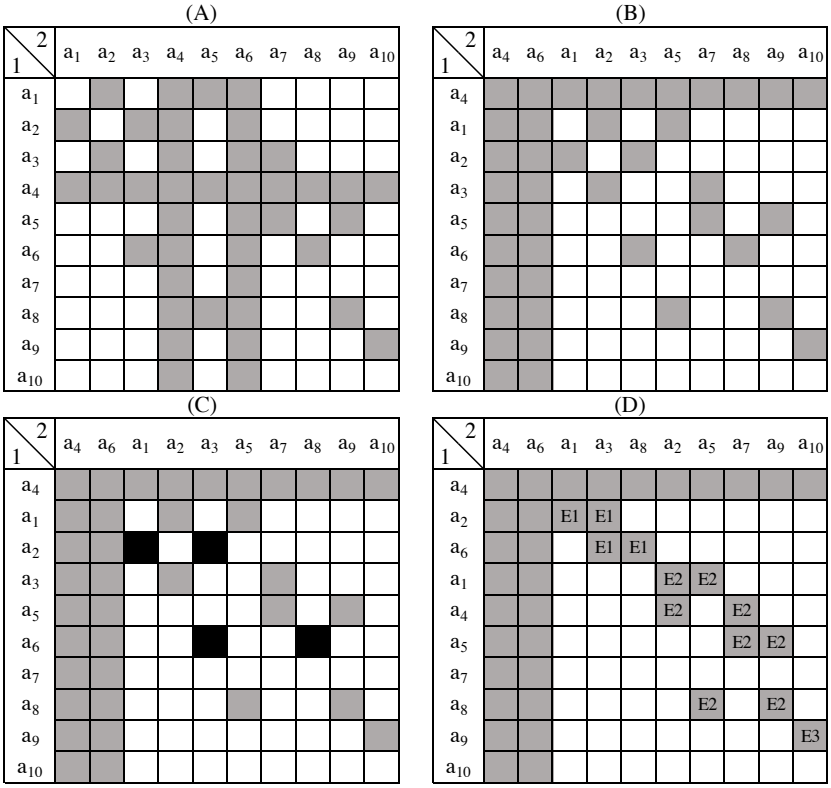


Figure 1.1: Two-agent Example

assigned a (local) dictator d_i so that for any $\succsim$ if $\tau(\succsim_1, \succsim_2) = (a, b)$

$$f(\succsim) = \begin{cases} (a,b) & \text{if } (a,b) \in C \\ (a, \max_{\succsim_2} C^2(a)) & \text{if } (a,b) \in E_k \text{ and } d_k = 1 \\ (\max_{\succsim_1} C^1(b), b) & \text{if } (a,b) \in E_k \text{ and } d_k = 2 \end{cases}$$

One can easily see that any local dictatorship is strategy-proof and Pareto efficient. The surprising fact is that the converse holds. That is, T directly indicates how to construct every mechanism.

Theorem 1. *$f : P^2 \to C$ is strategy-proof and Pareto efficient if and only if it is a local dictatorship.*

To see how this works for more familiar constraints, consider Figure 1.2. On the left is the house allocation constraint and on the right is the social choice constraint. Each grey square on the left is a different equivalence class of T, so every mechanism corresponds to a labeling of the grey boxes with 1's and 2's, which can be done independently. Another way to think about this is that each object is owned by one of the agents. If either agent top-ranks an object they own, they're guaranteed the ability to consume it. If both agents top-rank the other agents' object, they can trade. On the

right is the social choice constraint. Clearly T has a single block for this constraint since it is possible to move from any grey square to any other grey square, only changing one coordinate at a time, and only passing through grey squares. Then Theorem 1 immediately yields the two-agent version of the Gibbard–Satterthwaite Theorem, that every mechanism is a dictatorship. Famously, the Gibbard–Satterthwaite Theorem requires at least three alternatives. Our analysis provide a new perspective on this cardinality requirement: observe that if the social choice constraint in Figure 1.2 had only two objects, the constraint would be the top-left 2×2 constraint. In this case, T has two equivalence classes corresponding to the two grey squares.

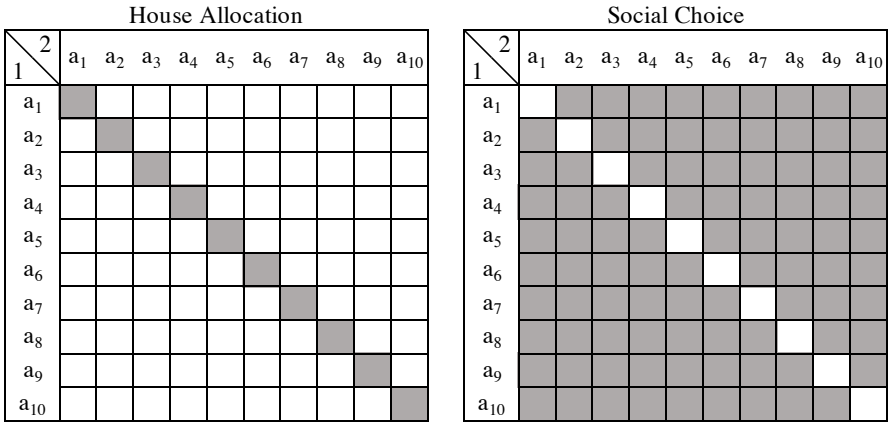


Figure 1.2: The social choice and house allocation constraints for two agents and 10 objects.

In independent and contemporaneous work, Meng (2019) provides an impressive characterization of all strategy-proof and Pareto efficient mechanisms for the two-agent social choice problem when agents are known to be indifferent between classes of alternatives that are fixed a priori. His characterization involves assigning a dictator at all profiles of preferences over announced indifference classes, where the dictator assignment must respect a cell-connected property. The structure of his result closely resembles our assignment of local dictators to the infeasible set. In fact, either result can be deduced from the other. However, these results are cast for very different questions, his for indifference and ours for constraints, so their substantive applications and contributions are quite different.

1.3.2 N Agents

When there are three or more agents, the approach we used for two agents fails to provide a straightforward characterization. The notion of a "local dictator" does not immediately generalize for more than two agents. One issue is that the set of compromising agents is not identified by knowing the local dictators because there are multiple agents besides the dictator. In fact, the ambiguity is deeper: not only is the identity

of the compromising agent ambiguous, but the number of compromising agents is not even necessarily fixed: it may be the case that having a single compromising agent is insufficient to move to a feasible solution, and instead multiple compromising agents must move to less-preferred assignments.

However, there is a subclass of constraints for which the basic intuition does follow the two-agents case and its characterization is therefore no more difficult. A constraint C is called **single-compromising** if for every infeasible allocation $(a_i)_{i\in N}$ and every agent i there is a reassignment a_i' for agent i such that (a_i', a_{-i}) is feasible. Thus, from any infeasible allocation, any agent can unilaterally compromise to make the social allocation feasible. In this case, every group strategy-proof and Pareto efficient mechanism can be written in a simple manner analogous to the characterization of the two-agent case. The generalization again partitions the space of infeasible allocations, but now each infeasible allocation is assigned a subset of agents who must compromise. We mention this special case where the two-agent approach extends because it exposes some of the limitations in generalizing that approach to more agents. First, it will be useful to have some definitions.

A **local compromiser assignment** is a map $\alpha : \mathcal{A} \to 2^N$ such that for every infeasible $x \in \bar{C}$, $\alpha(x)$ is nonempty and for every feasible $y \in C$, $\alpha(y) = \emptyset$. For $x \in \bar{C}$ an agent $i \in \alpha(x)$ is referred to as a **local compromiser** at x. This definition is motivated by the following algorithm, called the **constraint-traversing algorithm** for α, which take a profile of preferences as an input and returns a feasible allocation, or, if unable to do so, returns the symbol $\emptyset$. For a given preference profile $\succsim$:

Step 0 Let $x^0 = \tau_1(\succsim)$

Step k If x^{k-1} is feasible, stop and return x^{k-1}. Otherwise, if there is any $l \in \alpha(x^{k-1})$, such that $LC(x_l^{k-1})$ is empty, stop and return $\emptyset$. If not, define $x_i^k = x_i^{k-1}$ for all $i \notin \alpha(x^{k-1})$ and let $x_j^k = \max_{\succsim_j} LC(x_j^{k-1})$ for all $j \in \alpha(x^{k-1})$.

In words, the algorithm works by starting with the allocation in which all agents get their top choice. If this is feasible, the algorithm terminates. If not, there are number of local compromisers determined by α. The algorithm next tries the allocation in which the local compromisers switch to their next-best alternative, and the other agents keep their top choice. If this is feasible, the algorithm stops. Otherwise, there are again some local compromisers and the algorithm continues in the same manner. In this way the algorithm continues down agents' preference lists. For completeness, the statement of the algorithm includes a description of what to do if the algorithm exhausts an agents objects. The assumption that the constraint is single-compromising, along with proposition 2 will ensure that this never happens. When the constraint-traversing algorithm always yields a well-defined allocation, we call the induced mechanism a **constraint-traversing mechanism**. The following proposition gives a characterization of all group strategy-proof and Pareto efficient mechanisms for single-compromising constraints, analogous to Theorem 1 for the case with just two agents.

Proposition 2. *Let n be arbitrary and let C be single-compromising. A mechanism is group strategy-proof and Pareto efficient if and only if it is a constraint-traversing mechanism such that the local compromiser assignment satisfies*

1. $|\alpha(a)| \leq 1$ *for all* a
2. $\alpha(a) = i \implies \alpha(a'_i, a_{-i}) = i$ *whenever* $(a'_i, a_{-i}) \in \bar{C}$

An earlier working version of this paper included a more comprehensive examination of constraint-traversing mechanisms in general environments beyond single-compromising constraints, and this material is currently being incorporated into another paper.[20] For more general structures of constraints, constraint-traversing mechanisms are not necessarily incentive compatible and efficient, and the main work of this additional material is finding sufficient conditions that guarantee these conditions are satisfied.

From hereon, we consider the general case of arbitrary constraints, and not just single-compromising constraints. This will force the characterization to be more involved. For the remainder of this section, we will proceed with this characterization. The key insight is to consider marginal mechanisms, defined as follows.

Definition 2. Let $f : \mathscr{P} \to C$ and let M be a proper subset of N. Let $\succsim_{M^c}$ be a profile of preferences of agents not in M. The **marginal mechanism** of f holding M^c at $\succsim_{M^c}$ is denoted $f^M_{\succsim_{M^c}} : P^M \to \mathcal{O}^M$ and is defined by

$$\succsim \mapsto [f_j(\succsim, \succsim_{M^c})]_{j \in M}$$

we will denote $I^M(\succsim_{M^c}) = im(f^M_{\succsim_{M^c}})$ which will be referred to as M's **option set** holding M^c at $\succsim_{M^c}$

Thus a marginal mechanism holds fixed some of the agents' preferences $\succsim_{M^c}$ and defines an M-agent mechanism for the remaining agents, mapping their profile of announcements $\succsim_M$ to an M-agent allocation $f^M_{\succsim_{M^c}}(\succsim) \in \mathcal{O}^M$.

Clearly, marginal mechanisms inherit the group strategy-proofness of the original grand mechanism. The main result in this section shows that, going the other direction, it is enough to check that the two-agent marginal mechanisms are group strategy-proof to guarantee that the full mechanism is group strategy-proof.

Theorem 2. *The mechanism $f : \mathscr{P} \to C$ is group strategy-proof if and only if for every pair of agents $\{i, j\}$ and any profile $\succsim_{N \setminus \{i,j\}}$ of the other agents, the marginal mechanism of f holding $N \setminus \{i, j\}$ at $\succsim_{N \setminus \{i,j\}}$ is group strategy-proof.*

For two-agent mechanisms, there is only one group coalition—namely the grand coalition. Therefore group strategy-proofness of a two-agent mechanism is equivalent to individual strategy-proofness and Pareto efficiency on its image.

[20] Details are available from the authors upon request.

This drastically reduces the number of conditions one needs to check to ensure that a given mechanism is group strategy-proof. Rather than verifying incentives for all coalitions, it is sufficient to check that no two agents can profitably misreport their preferences. Furthermore, Theorem 2 is especially useful in conjunction with our previous characterization in Theorem 1 for all two-agent mechanisms. Application of Theorem 1 to all marginal mechanisms then provides a more explicit characterization of group strategy-proofness. We can show that the two-agent strategy-proof and Pareto efficient mechanisms form the "building blocks" of all group strategy-proof mechanisms. To do so we will need some notation. Let $F_n = \{f : P^n \to \mathcal{O}^n\}$ and $\mathscr{E}_{n,m} = \{\phi : P^n \to GS(\mathcal{O}^m)\}$. So $f \in F_n$ is just any map from the set of profiles for n agents to the set of possible allocations and any $\sigma \in \mathscr{E}_{n,m}$ provides, for each preference profile of n agents, a group strategy-proof mechanism for m other agents. Likewise, define $\mathscr{F}_{n,m} = \{\eta : P^n \to F_m\}$. We will need the following definition:

Definition 3. If $f \in F_n$ and $g \in F_m$ we may define the **direct sum** $f \oplus g : P^{n+m} \to \mathcal{O}^{n+m}$ by

$$f \oplus g(\succsim) = \left[f\left(\succsim_1, \succsim_2, \ldots, \succsim_n\right), g\left(\succsim_{n+1}, \succsim_{n+2}, \ldots, \succsim_{n+m}\right)\right]$$

This operation extends in the following way. For any $\sigma \in \mathscr{F}_{n,m}$ and $\rho \in \mathscr{F}_{m,n}$, we may define $\sigma \oplus \rho : P^{n+m} \to \mathcal{O}^{n+m}$ to be the map

$$\succsim \;\mapsto\; \left[\rho\left(\succsim_{n+1}, \cdots \succsim_{n+m}\right)\left(\succsim_1, \cdots \succsim_n\right), \sigma\left(\succsim_1, \cdots \succsim_n\right)\left(\succsim_{n+1}, \cdots \succsim_{n+m}\right)\right]$$

The final claim records these observations, explicitly providing a formula that characterizes the set of group strategy-proof mechanisms. This corollary says little other than 2, however explicitly justifies the notion the the two-agent mechanisms form the "building blocks" of arbitrary mechanisms.

Corollary 1.

$$GS(\mathcal{O}^n) = \bigcap_{\tau \in Sym(N)} \tau \circ \left[\mathscr{E}_{n-2,2} \oplus \mathscr{F}_{2,n-2}\right] \circ \tau^{-1}$$

Where $Sym(N)$ is the set of permutations of the agents N.

1.4 Applications

In this section, we will apply our general characterizations to specific constraints. These applications will feature a new class of mechanisms which are generalizations of serial dictatorships. In a basic serial dictatorship, agents take turns in a fixed order choosing their favorite objects among all objects which are feasible with the objects chosen by earlier dictators. In principle, the order of future agents might depend on earlier agents' choices. Our generalization of serial dictatorship does exactly that. We begin by formally describing the class of generalized serial dictatorships. We then apply this as well as our characterization results to the social choice problem and the roommates problem.

1.4.1 Generalized Serial Dictatorship

First, let us recall the definition of a serial dictatorship.

Definition 4. Let $\sigma(1), \ldots, \sigma(N)$ be a strict ordering of the agents $\{1, 2, \ldots, N\}$. For any constraint C, we may define the **serial dictatorship mechanism** which for each preference profile $\succsim$ gives the allocation defined by the following algorithm:

> Step 1 Agent $\sigma(1)$ chooses her favorite object a_1 from $\pi_{\sigma(1)}C$. Let μ_1 be the suballocation in which $\sigma(1)$ is assigned a_1 and all other agents are unassigned.
>
> Step k The agent $\sigma(k)$ chooses his favorite object a_k from $\pi_{\sigma(k)}C(\mu_{k-1})$. Let μ_k be the allocation whose graph is $G(\mu_{k-1}) \cup \{(\sigma(k), a_k)\}$. If all agents have been assigned an object, stop. If not, continue to step $k+1$.

Serial dictatorships are well-defined for any constraint and are always group strategy-proof and Pareto efficient.[21] It turns out, however, that we can easily generalize this notion to allow early dictators' choices to determine who will be the subsequent dictator. The main tension here is that, in order to maintain group strategy-proofness, we will have to ensure that the mechanism is nonbossy. That is, the early dictators will not be able to determine the subsequent order arbitrarily, but will be able to determine it only through the expression of their choices.

Recall that $\mathcal{S}$ is the set of suballocations (i.e. the maps $\mu : M \to \mathcal{O}$ where $M \subset N$). Let $\mathcal{S}'$ be the set of incomplete suballocations[22]. A **GSD-ordering** is a map $\zeta : \mathcal{S}' \to N$ such that for any suballocation μ, $\zeta(\mu)$ is an agent not allocated an object under μ. For each GSD-ordering and for any constraint C we may define a **generalized serial dictatorship mechanism** whose allocation at any preference profile is determined by the following algorithm:

> Step 1 The agent $d_1 \equiv \zeta(\emptyset)$ is the first dictator. She chooses her favorite object a_1 from $\pi_{d_1}C$. Let μ_1 be the suballocation in which d_1 is assigned a_1 and all other agents are unassigned.
>
> Step k The agent $d_k \equiv \zeta(\mu_{k-1})$ chooses her favorite object a_k from $\pi_{d_k}C(\mu_{k-1})$. Let μ_k be the allocation whose graph is $G(\mu_{k-1}) \cup \{(d_k, a_k)\}$. If all agents have been assigned an object, stop. If not, continue to step $k+1$.

Clearly, the standard serial dictatorship is the generalized serial dictatorship mechanism attained by setting $\zeta(\emptyset) = \sigma(1)$, $\zeta(\mu) = \sigma(2)$ for all suballocations μ in which a single agent is matched and so on. Unfortunately, a single mechanism can admit many GSD-orderings, that is, two different orderings might define the same mechanism.

[21] A fact we will prove shortly.

[22] M is a proper subset of N.

This is because the GSD-ordering ζ can be defined in any way off the "algorithm path" in the sense that, suballocations which will never be realized can be assigned any agent. For example, in the serial dictatorship mechanism, any allocation in which a single agent other than the dictator is assigned an object will never be realized, so the GSD assignment there is immaterial to the mechanism. Nevertheless, it is convenient to take $\mathcal{S}'$ as the domain of GSD-orderings. The following proposition shows that generalized serial dictatorships share the good incentive and efficiency properties of serial dictatorships.

Proposition 3. *For any constraint C, the generalized serial dictatorship mechanisms are group strategy-proof and Pareto efficient.*

Notice that this proposition demonstrates that $GS(C)$ is never empty.[23]

We can use these ideas to extend mechanisms defined on projections of the constraint. Suppose we have a constraint C and that for a proper subset $M \subset N$, we have a group strategy-proof and Pareto efficient mechanism f^M on the constraint C^M. Fix a GSD-ordering ζ. We will extend f^M to a mechanism on all of N and all of C by using a generalized serial dictatorship mechanism for agents in $N \setminus M$. In particular, define $(f^M, \zeta) : \mathscr{P} \to C$ via the following algorithm:

Step 1 Allocate $f_i^M(\succsim_M)$ to every agent i in M. Let μ_0 this suballocation. Let agent $d_1 = \zeta(\mu_0)$ choose her favorite object a_1 from among $\pi_{d_1} C(\mu_0)$ and let μ_1 be the suballocation whose graph is $G(\mu_0) \cup \{(d_1, a_1)\}$. If all agents have been allocated an object, stop. Otherwise, proceed to next step.

Step k The agent $d_k \equiv \zeta(\sigma_{k-1})$ chooses her favorite object x_k from $\pi_{d_k} C(\mu_{k-1})$. Let μ_k be the allocation whose graph is $G(\mu_{k-1}) \cup \{(d_k, x_k)\}$. If all agents have been assigned an object, stop. If not, continue to step $k+1$.

Proposition 4. *If $f^M : P^M \to C^M$ is Pareto efficient and group strategy-proof, for any GSD-ordering ζ, the mechanism (f^M, ζ) is group strategy-proof and Pareto efficient.*

1.4.2 The Roommates Problem

We now apply our general results to the canonical roommates problem. Our main contribution here is characterizing the group strategy-proof and Pareto efficient mechanisms for this problem.

In the roommates problem, an even number of agents who need to be paired as roommates. Each agent has a strict preference over the other agents as roommates. As discussed earlier, we can model this in our environment by letting $\mathcal{O} = N$ and using the constraint

$$C = \{\mu : N \to N \,|\, \mu(i) \neq i \text{ for all } i \text{ and } \mu^2 = id\}$$

[23] So long as the constraint is nonempty, which we assume throughout.

Any feasible mechanism for this constraint will be called a **roommates mechanism**. As mentioned in the introduction, the literature on the roommates problem has focused on the computational complexity of finding stable matching, and there is very little understanding of incentives and efficiency for one-sided matching.

Theorem 3 gives a full characterization of group strategy-proof and Pareto efficient mechanisms for the roommates problem. This is akin to the Gibbard–Satterthwaite Theorem that demonstrates all such mechanisms are dictatorships for the social choice problem and the recent result of Pycia and Ünver (2017) that characterizes all such mechanisms for the house allocation problem, but had not yet been discovered for one-sided matching. We settle this question for the roommates problem, and show that all mechanisms with these properties for the roommates problem are generalized serial dictatorships.

Theorem 3. *A roommates mechanism is group strategy-proof and Pareto efficient if and only if it is a generalized serial dictatorship.*

Although our results are generally unrelated to stability, this is one exception. As mentioned, a defining feature of the roommates problem is the lack of stable outcomes. One approach is to relax stability, with a possible direction to only require that pairs of agents where each ranks the other as her favorite must be matched. This weaker stability condition is called "mutually best" by Toda (2006) and "pairwise unanimity" by Takagi and Serizawa (2010). However, generalized serial dictatorships cannot satisfy even this very weak form of stability. So a corollary of Theorem 3 is that no group strategy-proof and Pareto efficient mechanism can satisfy mutual best or pairwise unanimity, exposing a tension between incentives and stability for the roommates problem. This negative observation for the roommates problem is not new; in fact, this corollary of our result can also be implicitly derived from Theorem 2 of Takamiya (2013) without an explicit characterization of group strategy-proofness.[24] Our constructive approach shows how this tension is related to the structure of the roommates problem as a constraint in our more general environment.

1.4.3 Social Choice

Here we apply the earlier theorems to provide a new proof for and insights into one of the canonical impossibility results of social choice,z by examining the structure of the social choice problem once it is expressed as a special constraint of our general model.

The first theorem in implementation theory was the celebrated negative result of Gibbard (1973) and Satterthwaite (1975) that the only strategy-proof and surjective social choice mechanisms are dictatorships. Since Pareto efficient mechanisms are necessarily surjective, this negative finding illuminates a fundamental tension between incentives and efficiency for social decisions. This tension can also be deduced as a corollary of our main result. Beyond providing a novel proof, our approach to the

[24]We thank Yuichiro Kamada for pointing this out to us.

Gibbard–Sattertwhaite Theorem yields additional insights that help understand the theorem more deeply. First, our environment for the theorem, in a model that includes social choice as a special case, demonstrates that the reason why social choice must yield a simple dictatorship, rather than a serial dictatorship, is because the structure of the constraint forces all agents' allocations to be immediately determined by fixing the dictator's allocation. If this feature is relaxed, then the dictator could consume her favorite object while still leaving flexibility in the allocation for other agents, that is, serial dictatorship is possible. So our approach shows how the dictatorship implied by the Gibbard–Satterthwaite Theorem can be seen as a special case of a more general feature of serial dictatorship.

Second and related, an immediate corollary of our main result is if all group strategy-proof mechanisms are serial dictatorships, then the marginal T relation, derived from the marginal constraint $C^{i,j}$, can have only one equivalence class. This provides a converse to the Gibbard–Sattertwhaite Theorem, showing that if all group strategy-proof mechanisms are serial dictatorships, then the constraint C must have a special structure. Again, this converse is only well-posed in a model where social choice is cast as a special case of private goods allocation, rather than vice versa as is more traditional.

One convenient feature of the diagonal social choice constraint is that, since all mechanisms are necessarily nonbossy to satisfy the constraint, there is no gap between group and individual strategy-proofness.[25]

Lemma 4. *Let C be the social choice constraint, i.e. $C = \{(a_i)_{i\in N} \mid a_i = a_j \text{ for all } i, j \in N\}$ then a map $f : \mathscr{P} \to C$ is group strategy-proof if and only if it is individually strategy-proof.*

We can then apply our main characterization results to the special case of the diagonal social choice constraint to derive that all group strategy-proof and onto mechanisms are dictatorships, which by virtue of Lemma 4 is equivalent to the Gibbard–Satterthwaite Theorem.

Theorem 4 (Gibbard–Satterthwaite)**.** *If $|\mathcal{O}| > 2$ and $f : \mathscr{P} \to C$ is surjective and strategy-proof then it is dictatorial.* [26]

As mentioned, the setup of our model enables us to sensibly ask the converse question: which types of constraints, beyond the diagonal social choice constraint, have the feature that all of the feasible, group strategy-proof mechanisms are (in some sense) dictatorial? In our context, the appropriate form of dictatorship is generalized serial dictatorship, since these always exist and specialize to dictatorship in the social choice setting. As a consequence of Proposition 4 and Theorem 1 we can show that if any

[25]This observation can also be alternatively deduced directly from the Gibbard–Sattertwhaite Theorem, since dictatorships are both individual and group strategy-proof. Since our aim is to prove that theorem, this is clearly not valid for our approach.

[26]In fact, we only need that $|im(f)| > 2$ in which case we could drop items never allowed and recover the same statement.

two-agent projection of the constraint is such that T has two equivalence classes, then $GS^N(C)$ admits mechanisms beyond GSD.

Theorem 5. *If a constraint C is such that for some i, j, the equivalence relation T on $C^{i,j}$ admits more than one equivalence class, $GS^n(C)$ is strictly larger than the set of generalized serial dictatorship mechanisms.*

1.5 Appendix

It will be convenient to introduce some additional notation for the proofs. If A and B are sets of objects and $\succsim \in P$, we say $A \succsim B$ if $a \succsim b$ for all $a \in A$ and $b \in B$. For disjoint sets of objects $A_1, A_2 \dots A_m$ we will denote

$$P[A_1, A_2 \dots A_m] = \{\succsim \in P \mid A_1 \succ A_2 \succ \cdots \succ A_m\}$$

and

$$P^{\uparrow}[A_1, A_2 \dots A_m] = \left\{ \succsim \in P \mid A_j \succ \mathcal{O} \setminus \bigcup_{i=1}^{j} A_i \text{ for all } j \right\}$$

When the A_i are singletons, we will abuse notation and drop the curly brackets, writing for example $P^{\uparrow}[a]$ to denote $P^{\uparrow}[\{a\}]$. We will also abuse notation slightly and use N to refer both to the set of agents and to the number of agents.

1.5.1 Proof of Propositon 1

We first need the following lemma, which is simply the forward direction of Lemma 3:

Lemma 5. *Let $f : \mathscr{P} \to \mathcal{A}$ be strategy-proof. Then for each i there is a nonempty correspondence $g_i : P^{n-1} \rightrightarrows \mathcal{O}$ such that for all $\succsim$*

$$f(\succsim) = \left(\max_{\succsim_i} g_i(\succsim_{-i}) \right)_{i \in N}$$

Proof. Define $g_i(\succsim_{-i}) = f_i(P, \succsim_{-i})$ then the result follows from strategy-proofness. □

We can now demonstrate the desired implications for the equivalence in turn:

(1) $\implies$ (2): Of course any group strategy-proof mechanism is individually strategy-proof. Suppose there is a profile $\succsim$ and an agent i with an alternative announcement $\succsim_i'$ such that $f_i(\succsim) = f_i(\succsim_i', \succsim_{-i})$ but for some j, $f_j(\succsim) \neq f_j(\succsim_i', \succsim_{-i})$. Then if $f_j(\succsim) \succ_j f_j(\succsim_i', \succsim_{-i})$, the coalition $\{i, j\}$ can improve their outcome at $(\succsim_i', \succsim_{-i})$ by announcing $(\succsim_i, \succsim_j)$. Conversely, if $f_j(\succsim) \prec_j f_j(\succsim_i', \succsim_{-i})$, the coalition $\{i, j\}$ can improve their outcome at $\succsim$ by announcing $(\succsim_i', \succsim_j)$.

(2) $\implies$ (3): Suppose we have two profiles $\succsim, \succsim' \in \mathscr{P}$ such that

$$LC_{\succsim_i'}[f_i(\succsim)] \supset LC_{\succsim_i}[f_i(\succsim)] \text{ for all } i$$

then notice that $f_1(\succsim_1', \succsim_2, \dots, \succsim_n) = f_1(\succsim)$ by Lemma 5 and by nonbossiness we have $f(\succsim_1', \succsim_2, \dots, \succsim_n) = f(\succsim)$. We can proceed, changing one preference at a time, to show that $f(\succsim') = f(\succsim)$ as desired.

(3) $\implies$ (1): Suppose f is Maskin monotonic; we will show that f is group strategy-proof. Let $\succsim \in \mathscr{P}$ and $\succsim_A'$ be a candidate violation for agents in A so that

$$f(\succsim_A', \succsim_{-A}) \succsim_j f(\succsim) \text{ for all } j \in A$$

we will show that this implies $f(\succsim'_A, \succsim_{-A}) = f(\succsim)$. For each $j \in A$ construct $\succsim^*_j$ to be identical to $\succsim_j$ except that it puts $f_j(\succsim'_A, \succsim_{-A})$ first. For any $j \in A$ we have

$$LC_{\succsim^*_j}(f_j(\succsim'_A, \succsim_{-A})) \supset LC_{\succsim_j}(f_j(\succsim'_A, \succsim_{-A})) \text{ and}$$
$$LC_{\succsim^*_j}(f_j(\succsim)) \supset LC_{\succsim_j}(f_j(\succsim))$$

for all j. The first is immediate. To see the second, notice that if $f_j(\succsim'_A, \succsim_{-A}) = f_j(\succsim)$ then it holds trivially. If instead, $f_j(\succsim'_A, \succsim_{-A}) \neq f_j(\succsim)$, by assumption we have $f_j(\succsim'_A, \succsim_{-A}) \succ_j f_j(\succsim)$ and since $\succsim^*$ only moves up the position of $f_j(\succsim'_A, \succsim_{-A})$, the second statement holds. However, by Maskin monotonicity, the first statement gives $f(\succsim^*_A, \succsim_{-A}) = f(\succsim'_A, \succsim_{-A})$ and the second gives $f(\succsim^*_A, \succsim_{-A}) = f(\succsim)$, so putting them together we get

$$f(\succsim'_A, \succsim_{-A}) = f(\succsim^*_A, \succsim_{-A}) = f(\succsim)$$

as desired. □

1.5.2 Proof of Lemma 1

By way of contradiction, suppose that $f : \mathscr{P} \to im(f)$ is group strategy-proof and that there is a profile $\succsim$ and an allocation $(a_i)_{i\in N} \in im(f)$ such that $a_i \succsim_i f_i(\succsim)$ for all i with at least one strict. By definition, there is an alternative profile $\succsim'$ such that $f(\succsim') = (a_i)_{i\in N}$ which is a profitable deviation from $\succsim$. □

1.5.3 Proof of Lemma 2

Let $\{g_i\}_{i\in N}$ be as in Lemma 3. For each j the preference $\succsim'_j$ does not change the relative ranking of the objects in $g_j(\succsim_{-j})$ hence we have $f_j(\succsim'_j, \succsim_{-j}) = f_j(\succsim)$ so by nonbossiness $f(\succsim'_j, \succsim_{-j}) = f(\succsim)$. Repeating this argument one agent at a time gives the result. □

1.5.4 Proof of Theorem 1 (Two-agent characterization)

($\Longleftarrow$) Applying lemma 3, we see that local dictatorships are strategy-proof. They are Pareto efficient by construction.
($\Longrightarrow$) If $C = \mathcal{O}^2$ then any Pareto efficient mechanism always gives both agents their top choice, which is trivially a local dictatorship.

Suppose now that C is a nonempty, proper subset of $\mathcal{O}^2$. By Lemma 2, it is without loss to assume that for any $(a, b) \in \bar{C}$ there are a' and b' with (a', b) and (a, b') in C. Fix $f : P^2 \to C$ which is strategy-proof and Pareto efficient.[27] The proof will proceed in two steps. First we show that for any infeasible allocation (a, b) there is a local dictator who gets their top choice at every preference profile where a and b are top-ranked respectively. Then we show that the local dictator is constant within blocks.

[27] Serial dictatorship always is both Pareto efficient and strategy-proof (as shown in proposition 3, so the set is nonempty.

Let $(a,b) \in \bar{C}$ and a', b' as above. Let $\succsim_1 \in P^{\uparrow}[a,a']$ and $\succsim_2 \in P^{\uparrow}[b,b']$. By Pareto efficiency, $f(\succsim_1,\succsim_2) = (a,b')$ or $f(\succsim_1,\succsim_2) = (a',b)$. Assume without loss that $f(\succsim_1,\succsim_2) = (a,b')$. We will show that this implies that 1 is the local dictator at (a,b). Pick any other $\succsim_2'$ which top-ranks b. By 2's strategy-proofness, $f_2(\succsim_1,\succsim_2') \neq b$, but then from Pareto efficiency, $f_1(\succsim_1,\succsim_2') = a$, since otherwise, the allocation (a',b) would Pareto dominate $f(\succsim_1,\succsim_2')$. Thus $f_1(\succsim_1,\succsim_2') = a$ whenever $\succsim_2' \in P^{\uparrow}[b]$. By 1's strategy-proofness, we have that $f_1(\succsim_1',\succsim_2') = a$ for all $\succsim_1', \succsim_2'$ with $\tau(\succsim_1',\succsim_2') = (a,b)$. Finally, by Pareto efficiency, $f(\succsim_1',\succsim_2') = (a, \max_{\succsim_2'} C^2(a))$ whenever $\tau(\succsim_1',\succsim_2') = (a,b)$. Thus we say that 1 is the local dictator at (a,b). Since (a,b), was arbitrary every other infeasible allocation has a local dictator.

Now suppose that $(a,b)B(a',b')$ and $(a,b) \neq (a',b')$. Then either $a = a'$ or $b = b'$. Without loss, assume $a = a'$. Suppose by way of contradiction that, that (a,b) and (a,b') have different local dictators. For example, suppose 1 is the local dictator at (a,b) and 2 is the local dictator at (a,b'). Consider the preference profile $(\succsim_1,\succsim_2)$ where $\succsim_1 \in P^{\uparrow}[a]$ and $\succsim_2 \in P^{\uparrow}[b,b',b'']$ where b'' is such that $(a,b'') \in C$. Then from the analysis above, we get $f(\succsim_1,\succsim_2) = (a,b'')$ since 1 is the local dictator at (a,b). However, if $\succsim_2' \in P^{\uparrow}[b']$, then $f_2(\succsim_1,\succsim_2') = b' \succ_2 b'' = f_2(\succsim_1,\succsim_2)$ since 2 is the local dictator at (a,b'), which is a violation of strategy-proofness. Thus either 1 is the local dictator at (a,b) and (a,b') or 2 is. For any two infeasible allocations (a,b) and (a',b') in an equivalence class of T, there is a sequence of infeasible allocations such that $(a,b)B(a_1,b_1)B\cdots B(a_n,b_n)B(a',b')$, so (a,b) and (a',b') have the same local dictator. □

1.5.5 Proof of Proposition 2

First we show that every group strategy-proof and Pareto efficient mechanism is constraint-traversing. Let C be a single-compromising constraint and fix and a group strategy-proof, Pareto efficient mechanism $f : \mathscr{P} \to C$. Let $a = (a_i)_{i\in N}$ be infeasible. For every i there is an object a_i' such that $(a_i', a_{-i}) \in C$. Let $\succsim_i \in P^{\uparrow}[a_i, a_i']$ for each i. Since f is feasible, there is at least one agent k who doesn't get their top choice at the constructed preference profile $\succsim = (\succsim_i)_{i\in N}$. However, Pareto-efficiency then implies that $f_i(\succsim) = a_i$ for all $i \neq k$ and $f_k(\succsim) = a_k'$. By Maskin monotonicity and Lemma 5 we have that for any $\succsim_{-k}'$ with $\max_{\succsim_j'} \mathcal{O} = a_j$ for all $j \neq k$, $a_k \notin g_k(\succsim_{-k})$, so that k always compromises when the top choice is a. Define $\alpha(a) = k$ (we can do this unambiguously because no other agent always compromises at a, e.g. at the profile $\succsim$). Since a was an arbitrary infeasible allocation, we can do the same for any other infeasible allocation to define α on all of $\bar{C}$. Finally, we establish inductively that f is constraint-traversing according to α. Pick any preference profile $\succsim'$. Start at $a^1 = (\max_{\succsim_i'} \mathcal{O})_{i\in N}$. If this is feasible, then f being Pareto efficient implies $f(\succsim') = a^1$. Otherwise, it is infeasible, and by the previous argument, we have an agent $k = \alpha(a^1)$ who must compromise. Replace $\succsim_k'$ with the same preference, except that it puts a_k^1 last. By Maskin monotonicity, this cannot affect the outcome of f. We therefore repeat the above process at the new profile. This is exactly how the constraint-traversing mechanism according to α works,

giving the result.

Now we need to show that α has to satisfy the property that if $\alpha(a) = i$ then for any $(a'_i, a_{-i}) \in \bar{C}$, we have $\alpha(a'_i, a_{-i}) = \{i\}$. However this follows from similar reasoning as in the two-agent case. If, instead $k = \alpha(a'_i, a_{-i})$ consider the profile $\succsim$ with $\tau(\succsim) = a$ and $\tau_2(\succsim_i) = a'_i$ and $\tau_2(\succsim_k) = a'_k$ where $(a'_k, a_{-k}) \in C$. We get a violation of Pareto efficiency since the constraint-traversing algorithm would make both i and k compromise to their second-best choice, which would be Pareto dominated by (a'_k, a_{-k}).

The fact that this mechanism is group strategy-proof and Pareto efficient is now a simple consequence of Maskin monotonicity and Proposition 1. □

1.5.6 Proof of Theorem 2 (N-agent characterization)

If f is group strategy-proof, the marginal mechanisms are group strategy-proof by definition. For the other direction, suppose that every two-agent marginal mechanism is group strategy-proof. Then f is individually strategy-proof since for any i and any profile $\succsim$ we can choose $j \neq i$ and consider the marginal mechanism $f^{i,j}_{\succsim_{-i,j}}$ then in this marginal mechanism i cannot profit from misreporting, hence she cannot in f. It remains to show that f is nonbossy. Now suppose we have $f_i(\succsim'_i, \succsim_{-i}) = f_i(\succsim)$ and for some j, $f_j(\succsim'_i, \succsim_{-i}) \neq f_j(\succsim)$, either $f_j(\succsim'_i, \succsim_{-i}) \succ_j f_j(\succsim)$ or $f_j(\succsim'_i, \succsim_{-i}) \prec_i f_j(\succsim)$. However, by assumption the marginal mechanism $f^{ij}_{\succsim_{-ij}}$ is group strategy-proof. From the two-agent characterization, no two-agent group strategy-proof mechanism can have this property. □

1.5.7 Proof of Corollary 1

The proof is an immediate application of Theorem 2. □

1.5.8 Proof of Proposition 3

Maskin monotonicity is easily seen to be satisfied, since starting from the first dictator, each agent will be given the same option set and will weakly prefer their original choice to any alternative. To see that it is Pareto efficient, by Lemma 1 it is enough to establish that its image is exactly C. By construction, the image is a subset of C. For any feasible allocation $a \in C$ let $\succsim_i$ put a_i first. Then $f(\succsim) = a$ so $im(f) = C$. □

1.5.9 Proof of Proposition 4

We will show that (f^M, ζ) is Maskin monotonic and Pareto efficient. Pick any $\succsim \in \mathscr{P}$ and let $\succsim'$ satisfy the conditions in the definition of Maskin monotonicity. I.e.

$$LC_{\succsim'_i}\left[(f^M, \zeta)_i(\succsim)\right] \supset LC_{\succsim_i}\left[(f^M, \zeta)_i(\succsim)\right] \text{ for all } i$$

Since f^M is group strategy-proof for the agents in M, it is Maskin monotonic. Hence we have $f^M(\succsim_M) = f^M(\succsim'_M)$, then by definition, $(f^M, \zeta)_i(\succsim') = (f^M, \zeta)_i(\succsim)$ for all

$i \in M$. As a consequence, the sequence of dictators is the same. Thus we have Maskin monotonicity.

By Lemma 1 it is enough to establish that the image of (f^M, ζ) is exactly C. To see this, let $(a_i)_{i\in N} \in C$, since f^M is Pareto efficient on C^M there is some profile $\succsim_M$ with $f^M(\succsim_M) = (a_i)_{i\in M}$. For agents not in M let $\succsim_j \in P^{\uparrow}(a_j)$. At this profile, we have $(f^M, \zeta) = (a_i)_{i\in N}$ as desired. □

1.5.10 Proof of Theorem 3 (Roommates characterization)

The "if" direction follows directly from Proposition 3.

We will prove the "only if" Theorem by mathematical induction. First, by Lemma 2, we ignore any agents' ranking for infeasibly matching with herself. If $N = 2$ there is only one feasible allocation, so every mechanism is trivially a generalized serial dictatorship. If $N = 4$, then the problem is a social choice problem since a single agent's match determines the full outcome. In this case, the result follows from the Gibbard–Satterthwaite Theorem. Suppose that for all $m < n$ when there are $2m$ agents, all group strategy-proof and Pareto efficient roommates mechanisms are generalized serial dictatorships. We will show this for $2n$ agents. It will be enough to show that there is an agent j such that $f_j(\succsim) = \max_{\succsim_j} N$ for all $\succsim$, since, conditional on each of j's choices, the remaining $2n - 2$ agents need to assigned a roommate, which itself gives a roommates mechanism guaranteed to be a generalized serial dictatorship by the induction assumption.

Let f be a group strategy-proof and Pareto efficient roommates mechanism for $2n$ agents with $n \geq 3$. We will first consider the possible two-agent marginal mechanisms. Let $i \neq j$ and fix a profile $\succsim_{-ij}$ of the other agents. Assume $(j, i) \in I^{ij}(\succsim_{-ij})$, so that it is possible for i and j to match when the other agents announce $\succsim_{-ij}$. For all $k \neq i$, $(j, k) \notin I^{ij}(\succsim_{-ij})$ since (j, k) has i matched to j but j matched to k. Likewise, for all $k \neq j$ we have $(k, i) \notin I^{ij}(\succsim_{-ij})$. Define $R_i = \{x \in N \,|\, (x, y) \notin I^{ij}(\succsim_{-ij}) \text{ for all } y \in N\}$ and $R_j = \{y \in N \,|\, (x, y) \notin I^{ij}(\succsim_{-ij}) \text{ for all } x \in N\}$. Then after possibly permuting agents, we get a marginal constraint like the one shown on the left of Figure 1.3, with the exception that some non-grey squares on the bottom right may actually be infeasible. As usual, we will ignore agents preferences over objects they can never receive [28]. If $[N - R_i \cup \{j\}] \times [N - R_j \cup \{i\}]$ intersects any infeasible point, then the equivalence relation T has a single equivalence class, as illustrated on the right-hand picture of Figure 1.3.[29] Therefore there must be a single dictator in the marginal mechanism $f^{ij}_{\succsim_{-ij}}$ by Theorems 1 and 2. Otherwise, every allocation in $[N - R_i \cup \{j\}] \times [N - R_j \cup \{i\}]$ is feasible or the set is empty. In the latter case $I^{ij}(\succsim_{-ij})$ is a singleton, and obviously only one marginal mechanism. In the former case, as a consequence of Theorem 1 there are four possible Pareto efficient, strategy-proof marginal mechanisms as illustrated in Figure 1.4.

[28] In this case, R_i and R_j are not possible for i and j to match holding fixed the preferences $\succsim_{-j}$

[29] Recall the relation T was defined immediately before the statement of Theorem 1.

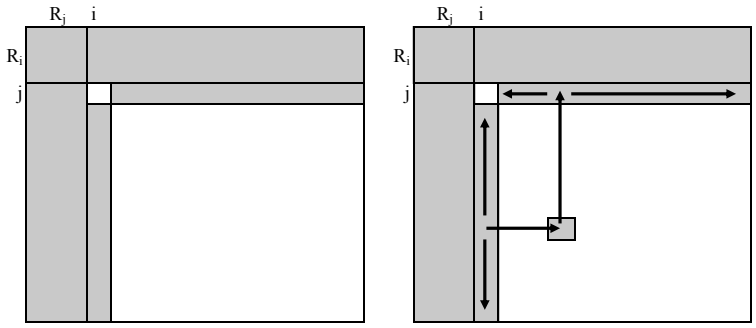


Figure 1.3: The marginal Constraint $I^{ij}(\succsim_{-ij})$

In panel (A), j is the dictator since i must compromise at every infeasible allocation. In panel (B), i is the dictator. In Panel (C), i and j are matched together if either top-ranks the other and are only unmatched if both i prefers someone in $N - R_i \cup \{j\}$ and j prefers someone in $N - R_j \cup \{i\}$. In panel (D), i and j are matched only if both top-rank the other and are unmatched otherwise.

Summarizing, if $(j, i) \in I^{ij}(\succsim_{-ij})$, there are four possible types of mechanisms $f^{ij}_{\succsim_{-ij}}$:

1. $f^{ij}_{\succsim_{-ij}}$ is constant and (j, i). In this case, $N - R_i = \{j\}$ and $N - R_j = \{i\}$.

2. $f^{ij}_{\succsim_{-ij}}$ is dictatorial, so i gets their top choice from $N - R_i$ or j gets their top choice from or $N - R_j$ and the other agent gets their top choice consistent with the dictators' allocation. Note that in a dictatorial mechanism, the non-dictator cannot affect the option set of the dictator.

3. i and j are matched by default, and are unmatched only if both agree. This is shown in panel (C). In this case, all allocations in $[N - R_i \cup \{j\}] \times [N - R_j \cup \{i\}]$ are feasible.

4. i and j are unmatched by default and are matched only if both agree. This is shown in Panel (D). In this case, all allocations in $[N - R_i \cup \{j\}] \times [N - R_j \cup \{i\}]$ are feasible.

In the remainder of the proof, we will often need to show that a given two-agent marginal mechanism is dictatorial. To do that, we need show that it is possible for both agents to match with one another, that it is non-constant (i.e. that there are at least two possible allocations for the two agents holding the other agents' preferences fixed), and that it is not of the third or fourth types. The third type of mechanism is usually easy to rule out. If we can find a preference where one agent top-ranks the other and they are still not matched, it cannot be of type three. Type (4) is somewhat more subtle, but we can rule it out if an agent can match with a second agent even when that agent bottom-ranks the first agent.

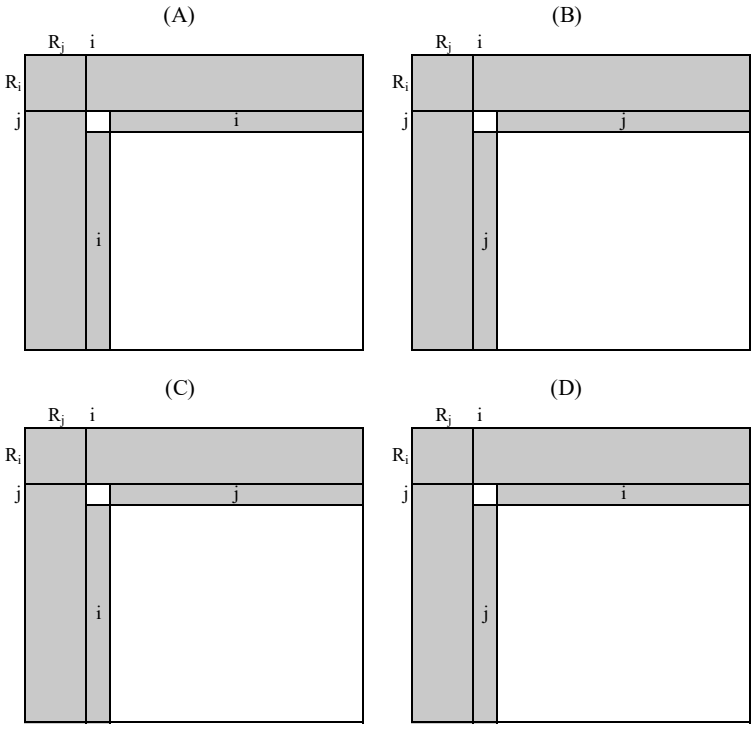


Figure 1.4: The possible marginal mechanisms $f^{ij}_{\succsim_{-ij}}$

We now need the following lemma, whose validity depends on the induction assumption.

Lemma 6. *Let A be a nonempty proper subset of N with an even number of agents and $|A| \geq 4$. If $\succsim_{N\setminus A} \in \left[P^{\uparrow}(N \setminus A)\right]^{N\setminus A}$, then there is an agent $j \in A$ such that*

$$f_j(\succsim_A, \succsim_{N\setminus A}) = \max_{\succsim_j} N$$

whenever $\max_{\succsim_j} N \in A$. *Equivalently,* $g_j(\succsim_{A-\{j\}}, \succsim_{N\setminus A}) \supset A - \{j\}$ *for all* $\succsim_{A-\{j\}}$.

Proof. For notational convenience, let $A = \{1, 2, \ldots, l\}$ and $N \setminus A = \{l+1, \ldots N\}$. Fix a profile $\succsim_{N\setminus A}$. For any $\succsim_1, \cdots \succsim_l \in P^{\uparrow}(\{1, 2, \ldots, l\})$, by Pareto efficiency, $f(\succsim_1, \succsim_2, \ldots, \succsim_l, \succsim_{N\setminus A})$ will match agents in $\{1, 2, \ldots, l\}$ with other agents in $\{1, 2, \ldots, l\}$ and agents in $\{l+1, \ldots N\}$ with other agents in $\{l+1, \ldots N\}$. Thus the marginal mechanism $f(\cdot, \succsim_{N\setminus A})$ restricted to profiles in $\left[P^{\uparrow}(\{1, 2, \ldots, l\})\right]^l$ gives a roommates mechanism for the agents in $\{1, 2, \ldots, l\}$. By the group strategy-proofness and efficiency of f, the marginal mechanism is also group strategy-proof and efficient. By the induction assumption this marginal mechanism is a generalized serial dictatorship. Without loss, assume that 1 is the first dictator. Then we have $g_1(\succsim_2, \ldots, \succsim_l, \succsim_{N\setminus A}) \supset \{2, 3, \ldots l\}$ for all $\succsim_2, \ldots, \succsim_l$ in $P^{\uparrow}(\{1, 2, \ldots, l\})$. For any $\succsim_3, \ldots, \succsim_l$ in $P^{\uparrow}(\{1, 2, \ldots, l\})$, consider the $1, 2$-marginal mechanism. Since $g_1(\succsim_2, \ldots, \succsim_l, \succsim_{N\setminus A}) \supset \{2.3. \ldots l\}$ for all $\succsim_2, \ldots, \succsim_l$

in $P^{\uparrow}(\{1,2,\ldots,l\})$, if 1 top-ranks 2 and 2 announces a preference in $P^{\uparrow}(\{1,2,\ldots,l\})$, 1 and 2 are matched. Thus $(2,1) \in I^{1,2}(\succsim_3,\ldots,\succsim_l,\succsim_{N\setminus A})$. From the considerations above, there are four possibilities for this marginal mechanism. Let $\succsim_1^*$ top rank $j \neq 2$ and $j \leq l$ and $\succsim_2$ in $P^{\uparrow}(\{1,2,\ldots,l\})$ top-rank 1. At this profile, 1 and j are matched. Hence the $1,2$ marginal mechanism is not constant. Furthermore, it cannot be of type (3), since 1 is matched with j, despite 2 top-ranking 1. Let $\succsim_2^*$ be in $P^{\uparrow}(\{1,2,\ldots,l\})$ and top-rank her match at the profile $(\succsim_1^*,\succsim_2)$. Since 1 and 2 are matched when 1 top-ranks 2 and 2 announces $\succsim_2^*$, the mechanism also cannot be of type (4) (At $\succsim_2^*$, agent 2 is top-ranking a feasible match in the $1,2$ marginal mechanism, but 1 can still match with her). The only possibility left is that the $1,2$-marginal mechanism is dictatorial with agent 1 as the dictator. Since non-dictators cannot affect the option set of dictators, we get that $g_1(\succsim_2',\succsim_3,\ldots,\succsim_l,\succsim_{N\setminus A}) \supset \{2.3.\ldots l\}$ for any $\succsim_2'$ and any $\succsim_3,\ldots,\succsim_l$ in $P^{\uparrow}(\{1,2,\ldots,l\})$. We could have carried out the above argument with any i in place of 2, so in fact we have

$$g_1(\succsim_2,\ldots,\succsim_{i-1},\succsim_i',\succsim_{i+1},\cdots\succsim_l,\succsim_{N\setminus A}) \supset \{2,3,\ldots l\}$$

for any $\succsim_i'$ and any $\succsim_2,\ldots,\succsim_{i-1},\succsim_{i+1},\cdots\succsim_l$ in $P^{\uparrow}(\{1,2,\ldots,l\})$.

The goal is to show that

$$g_1(\succsim_2',\ldots,\succsim_l,\succsim_{N\setminus A}) \supset \{2,3,\ldots l\}$$

for all $\succsim_2',\ldots,\succsim_l'$. We will do this by induction. Specifically we will show that if for any $0 < q-1 < l-1$ and any $A' \subset A-\{1\}$ with $|A'| = q-1$ we have $g_1(\succsim_{A'}',\succsim_{A-A'\cup\{1\}},\succsim_{N\setminus A}) \supset \{2,3,\ldots l\}$ for any $\succsim_{A'}'$ and any $\succsim_{A-A'\cup\{1\}}$ in $\left[P^{\uparrow}(A)\right]^{A-A'\cup\{1\}}$ then the same holds for any $A' \subset A-\{1\}$ with q agents.

For simplicity, let $A' = \{2,\ldots q+1\}$ and pick any $\succsim_2',\ldots,\succsim_{q+1}'$. By the induction assumption, we have $g_1(\succsim_2',\ldots,\succsim_q',\succsim_{q+1},\cdots\succsim_l,\succsim_{N\setminus A}) \supset \{2,3,\ldots l\}$ for any $\succsim_2',\ldots,\succsim_q'$ and any $\succsim_{q+1},\cdots\succsim_l$ in $P^{\uparrow}(A)$. Now by the same arguments as above, the $1,q+1$-marginal mechanism at this profile is either of type (2) (i.e. dictatorial) or it is of type (4). Suppose, by way of contradiction, that it is of type (4) and let $\succsim_{q+1}^*$ bottom-rank 1. Then doing so removes $q+1$ from 1's option set, but leaves it otherwise the same. Let $\succsim_1^{**}$ top-rank $q+1$ and second-rank q. From the above discussion, we get that 1 is matched to q at the marginal profile $(\succsim_1^{**},\succsim_{q+1}^*)$. If we let $\succsim_q^* \in P^{\uparrow}(A)$ top-rank 1, then by Maskin-monotonicity, we have

$$f(\succsim_1^{**},\succsim_2',\ldots,\succsim_q',\succsim_{q+1}^*,\succsim_{q+2},\ldots,\succsim_l,\succsim_{N\setminus A}) = f(\succsim_1^{**},\succsim_2',\ldots,\succsim_{q-1}',\succsim_q^*,\succsim_{q+1}^*,\succsim_{q+2},\ldots,\succsim_l,\succsim_{N\setminus A})$$

but on the left we have 1 is matched to q, her second-top choice. By the induction assumption, on the right we should have $q+1$ in 1's option set since the agents $q, q+2,\ldots,l$ are all announcing a preference in $P^{\uparrow}(A)$, leaving only $q-1$ agents announcing a possibly different preference. This gives a contradiction so we must have that 1 is the dictator in the $1,q+1$-marginal mechanism. □

We will call agent j in the lemma above, the *marginal dictator*. Having done this, the idea is to partition the agents in two ways. First we consider the partition $\{1,2\}\{3,4,\ldots,N\}$. By lemma 6, for $\succsim_1^* \in P^\uparrow(2)$ and $\succsim_2^* \in P^\uparrow(1)$ there is a marginal dictator among $\{3,4,\ldots,N\}$. Second, we consider the partition $\{1,2,3,4\},\{5,6,\ldots,N\}$ and again lemma 6 says that given $\succsim_5^*,\ldots,\succsim_n^* \in P^\uparrow(\{5,\ldots,n\})$, there is a marginal dictator among $\{1,2,3,4\}$. We show that by comparing these two dictators, we can find a single dictator for the whole mechanism.

As above, let $\succsim_1^* \in P^\uparrow(2)$, $\succsim_2^* \in P^\uparrow(1)$ and without loss assume that 3 is the marginal dictator among $\{3,\ldots N\}$. By Maskin-monotonicity, it is also without loss to suppose that both $\succsim_1^*$ and $\succsim_2^*$ bottom-rank 3 [30]. Also choose $\succsim_5^*,\ldots,\succsim_n^* \in P^\uparrow(\{5,\ldots,n\})$. By lemma 6 $g_3(\succsim_1^*,\succsim_2^*,\succsim_4',\ldots,\succsim_N') \supset \{4,\ldots,N\}$ for all $\succsim_4',\ldots,\succsim_N'$. Likewise, for some $i \in \{1,2,3,4\}$, we have $g_i(\succsim'_{\{1,2,3,4\}-\{i\}},\succsim_5^*,\ldots,\succsim_n^*) \supset \{1,2,3,4\}-\{i\}$ for all $\succsim'_{\{1,2,3,4\}-\{i\}}$. This gives four cases, corresponding to the possible identities of i. However, note that i cannot be 4 since 3 and 4 are matched at the profile $(\succsim_1^*,\succsim_2^*,\succsim_3,\succsim_4,\succsim_5^*,\ldots,\succsim_n^*)$ where 3 top ranks 4 regardless of $\succsim_4$. Since 1 and 2 are so far symmetric, this leaves two cases: $i=1$ (and $i=2$ by symmetry) and $i=3$.

We will start with the latter case. So we have

$$g_3(\succsim_1^*,\succsim_2^*,\succsim_4',\ldots,\succsim_N') \supset \{4,\ldots,N\} \text{ for all } \succsim_4',\ldots,\succsim_N' \text{ , and} \tag{1.1}$$

$$g_3(\succsim_1',\succsim_2',\succsim_4',\succsim_5^*,\ldots,\succsim_n^*) \supset \{1,2,4\} \text{ for all } \succsim_1',\succsim_2',\succsim_4' \tag{1.2}$$

In particular, $g_3(\succsim_1^*,\succsim_2^*,\succsim_4',\succsim_5^* \ldots,\succsim_N^*) = N-\{3\}$ for all $\succsim_4'$. Consider the $3,5$-marginal mechanism at the profile $\succsim_1^*,\succsim_2^*,\succsim_4',\succsim_6^* \ldots,\succsim_N^*$ for any $\succsim_4'$. From equation 1.1 above, 3 and 5 are matched whenever 3 top ranks 5, regardless of 5's preference. It is also possible for 3 to match with 4 regardless of 5's preference. From the discussion about the possible two-agent marginal mechanisms, the only possibility for this marginal mechanism has 3 as the dictator. In this case, 5's announcement cannot affect 3's option set. Thus we have $g_3(\succsim_1^*,\succsim_2^*,\succsim_4',\succsim_5',\succsim_6^* \ldots,\succsim_N^*) = N-\{3\}$ for any $\succsim_4',\succsim_5'$. Repeating this argument one agent at a time implies that

$$g_3(\succsim_1^*,\succsim_2^*,\succsim_4',\ldots,\succsim_N') = N-\{3\} \text{ for all } \succsim_4',\ldots,\succsim_N' \text{ , and} \tag{1.3}$$

a symmetric argument shows that

$$g_3(\succsim_1',\succsim_2',\succsim_4',\succsim_5^*,\ldots,\succsim_n^*) = N-\{3\} \text{ for all } \succsim_1',\succsim_2',\succsim_4' \text{ .} \tag{1.4}$$

Let $\succsim_1^{**}$ be identical to $\succsim_1^*$, except that 3 is top ranked. Define $\succsim_2^{**}$ equivalently. Now we want to show that the following three equations hold:

$$g_3(\succsim_1^{**},\succsim_2^*,\succsim_4',\ldots,\succsim_N') = N-\{3\} \text{ for all } \succsim_4',\ldots,\succsim_N' \text{ , and} \tag{1.5}$$

[30] Let $\succsim_1^* \in P^\uparrow(2)$, $\succsim_2^* \in P^\uparrow(1)$, by lemma 6, we have $g_3(\succsim_1^*,\succsim_2^*,\succsim_4',\ldots,\succsim_N') \supset \{4,\ldots,N\}$ Let $\succsim_1^{**}$ and $\succsim_2^{**}$ be the same as $\succsim_1^*$ and $\succsim_2^*$ respectively, except both bottom-rank 3. Let $\succsim_3$ top rank $k \in \{4,\ldots,N\}$. Then $f_3(\succsim_1^*,\succsim_2^*,\succsim_3,\succsim_4',\ldots,\succsim_N') = k$ for any $\succsim_4',\ldots,\succsim_N'$. But Maskin-monotonicity then says $f_3(\succsim_1^{**},\succsim_2^{**},\succsim_3,\succsim_4',\ldots,\succsim_N') = k$ for any $\succsim_4',\ldots,\succsim_N'$.

$$g_3(\succsim_1^*, \succsim_2^{**}, \succsim_4', \ldots, \succsim_N') = N - \{3\} \text{ for all } \succsim_4', \ldots, \succsim_N' \text{ , and} \tag{1.6}$$

$$g_3(\succsim_1^{**}, \succsim_2^{**}, \succsim_4', \ldots, \succsim_N') = N - \{3\} \text{ for all } \succsim_4', \ldots, \succsim_N' \text{ .} \tag{1.7}$$

Since the arguments are all symmetric, we will just show equation 1.5. From equation 1.4, we know that $g_3(\succsim_1^{**}, \succsim_2^*, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*) = N - \{3\}$. Consider the $3,5$-marginal mechanism at the profile $\succsim_1^{**}, \succsim_2^*, \succsim_4', \succsim_6^*, \ldots, \succsim_N^*$. Let $\succsim_3 \in P^\uparrow(5,4)$ and $\succsim_3'' \in P^\uparrow(4,5)$. Then we have that 3 and 5 are matched at the profile $(\succsim_3, \succsim_5^*)$ and 3 and 4 are matched at the profile $(\succsim_3'', \succsim_5^*)$. Thus the $3,5$-marginal mechanism is not constant and if it is dictatorial, 3 must be the dictator. We must also rule out type (3) and type (4) mechanisms. Let $\succsim_5''$ top-rank 3. In a type (3) mechanism, we would have that 3 and 5 are matched at the profile $(\succsim_3'', \succsim_5'')$. However, going back to the full mechanism, this would imply, by Maskin monotonicity that

$$f(\succsim_1^{**}, \succsim_2^*, \succsim_3'', \succsim_4', \succsim_5'', \succsim_6^*, \ldots, \succsim_N^*) = f(\succsim_1^*, \succsim_2^*, \succsim_3'', \succsim_4', \succsim_5'', \succsim_6^*, \ldots, \succsim_N^*)$$

however, on the right hand side, we have 3 matched with 4 by equation 1.3. Thus the $3,5$-marginal mechanism cannot be of type (3). Finally, suppose that $\succsim_5'''$ ranks agent 3 last. If the marginal mechanism were type (4), we could not have 3 and 5 matched at $(\succsim_3, \succsim_5''')$. However, in a type (4) mechanism, either agent can only remove themselves from the other agents option set. Hence in this case we would have that 3 is matched to 4 at $(\succsim_3, \succsim_5''')$. But for the same reasons as above, Maskin monotonicity implies this cannot happen. Hence 3 is the dictator in the marginal mechanism and 5's preference does not affect 3's option set so

$$g_3(\succsim_1^{**}, \succsim_2^*, \succsim_4', \succsim_5', \succsim_6^*, \ldots, \succsim_N *) = N - \{3\} \text{ for all } \succsim_4', \succsim_5' \text{ .}$$

Repeating this argument one agent at a time gives us equation 1.5.

Now we claim that equations 1.3 and 1.6, together imply that

$$g_3(\succsim_1^*, \succsim_2', \succsim_4', \ldots, \succsim_N') = N - \{3\} \text{ for all } \succsim_2', \succsim_4', \ldots, \succsim_N' \tag{1.8}$$

Equation 1.3 says that 3 has the option to match with 2, even though 2 bottom-ranks 3 by assumption. Equation 1.6 that 3 has the option to not match with 2, even if 2 top ranks her. Thus we can only have 3 as the marginal dictator in the $2,3$-marginal mechanism at any $\succsim_1^*, \succsim_4', \ldots, \succsim_N'$. Since 2 cannot affect 3's option set, we get eqation 1.8. Repeating the same arguments with equations 1.5 and 1.7 show that

$$g_3(\succsim_1^{**}, \succsim_2', \succsim_4', \ldots, \succsim_N') = N - \{3\} \text{ for all } \succsim_2', \succsim_4', \ldots, \succsim_N' \tag{1.9}$$

Finally, by comparing equations 1.8 and 1.9, we get the desired result that $g_3(\succsim_1', \succsim_2'$ $, \succsim_4', \ldots, \succsim_N') = N - \{3\}$ for all $\succsim_2', \succsim_4', \ldots, \succsim_N'$.

Now we must come to the case in which 1 is the marginal dictator among $\{1,2,3,4\}$ at the profile $\succsim_5^*, \ldots, \succsim_N^*$. Our strategy will be to reduce this to the previous case by showing that for some $\succsim_3^\dagger \in P^\uparrow(4)$, $\succsim_4^\dagger \in P^\uparrow(3)$, that 1 is also the marginal dictator among $\{1,2,5,\ldots,N\}$.

By lemma 6, we have

$$g_1(\succsim_2', \succsim_3', \succsim_4', \succsim_5^*, \ldots, \succsim_n^*) \supset \{2,3,4\} \text{ for all } \succsim_2', \succsim_3', \succsim_4' \tag{1.10}$$

Let $k \in 5, \ldots, N$ and $\succsim_3$ top-rank k. Then f matches 1 and 2 and also 3 and k in the match $f(\succsim_1^*, \succsim_2^*, \succsim_3, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*)$ for any $\succsim_4' \in P^{\uparrow}(\{3, \ldots, N\})$. Let $\succsim_2^{**}$ be the same as $\succsim_2^*$, except that it top-ranks 3 and let $\succsim_3^{**}$ be the same as $\succsim_3$, except that it top-ranks 2. Since 1 is the marginal dictator among $\{1,2,3,4\}$, 1 and 2 are still matched at the profile $(\succsim_1^*, \succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*)$, so by Maskin monotonicity, we have

$$f(\succsim_1^*, \succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*) = f(\succsim_1^*, \succsim_2^*, \succsim_3^*, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*)$$

and in particular, 3 and k are still matched. Now consider the $1,k$-marginal mechanism at $(\succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_{k+1}^*, \ldots, \succsim_N^*)$. Let $\succsim_1^{**}$ be the same as $\succsim_1^*$, except that it top-ranks k and let $\succsim_k^{**}$ be the same as $\succsim_k^*$, except that it top-ranks 1. We must have that 1 and k are matched in the marginal mechanism at $(\succsim_1^{**}, \succsim_k^{**})$, since otherwise Maskin monotonicity says that f gives the same result as though they had announced $(\succsim_1^*, \succsim_k^*)$, but in this case, 1 and 2 are matched and 3 and k are matched which is inefficient since we could swap 1 and 3's matches. Thus $(k,1)$ is in $I^{1,k}(\succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_{k+1}^*, \ldots, \succsim_N^*)$. From the considerations above, there are four possibilities for this mechanism. However, since both $(2,3)$ and $(k,1)$ are in the marginal option set, the marginal mechanism is not constant. Note also that if 1 top ranks 3 and k announces $\succsim_k^*$, then by equation 1.10, 1 and 3 are matched. Thus it is possible for both 1 and k to match with 3 in this marginal mechanism. But since both can't match with 3 at the same time, the marginal constraint is like the one shown on the right of figure 1.3, and there must be a single dictator. We will show that this dictator must be 1. To do this, we will have to take a detour to the $3,l$-marginal mechanism.

By equation 1.10, $f_1(\succsim_1^*, \succsim_2^{**}, \succsim_3', \succsim_4', \succsim_5^*, \ldots, \succsim_N^*) = 2$ for all $\succsim_3', \succsim_4'$, so by Maskin monotonicity, we have

$$f(\succsim_1^*, \succsim_2^{**}, \succsim_3', \succsim_4', \succsim_5^*, \ldots, \succsim_N^*) = f(\succsim_1^*, \succsim_2^*, \succsim_3', \succsim_4', \succsim_5^*, \ldots, \succsim_N^*)$$

for all $\succsim_3', \succsim_4'$. In particular, we have $g_3(\succsim_1^*, \succsim_2^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*) = \{4, \ldots, N\}$ for all $\succsim_4'$ by equation 1.3. Consider the $3,k$-marginal mechanism at this profile. If 3 top ranks k they are matched. If 3 top ranks 4 they are not. In the latter case, k is matched to someone from $\{5, \ldots, N\}$, which she prefers. Hence the marginal mechanism is either a dictatorship with 3 as the dictator, or it is of the third type in which 3 and k are matched if either top-ranks the other. Let $\succsim_3''$ top rank 4 and $\succsim_k''$ top rank 3. In the type (3) marginal mechanism, we would have 3 and k matched in

$$f(\succsim_1^*, \succsim_2^{**}, \succsim_3'', \succsim_4', \succsim_5^*, \cdots \succsim_{k-1}^*, \succsim_k'', \succsim_{k+1}^*, \ldots, \succsim_N^*)$$

but then Maskin-monotonicity would imply that we get the same outcome if 2 announced $\succsim_2^*$, yet at this profile , by equation 1.3, we would have 3 matched to 4.

Hence we have that 3 is the dictator in the $3,k$-marginal mechanism at $(\succsim_1^*, \succsim_2^{**}, \succsim_4'$ $, \succsim_5^*, \cdots \succsim_{k-1}^*, \succsim_{k+1}^*, \ldots, \succsim_N^*)$ for all $\succsim_4'$. This implies that $g_3(\succsim_1^*, \succsim_2^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*$ $, \succsim_k', \succsim_{k+1}^*, \ldots, \succsim_N^*) = \{4, \ldots, N\}$ for all $\succsim_4'$ and $\succsim_k'$. So we have $f_3(\succsim_1^*, \succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*$ $, \ldots, \succsim_{k-1}^*, \succsim_k^{**}, \succsim_{k+1}^*, \ldots, \succsim_N^*) = k$, and by non-bosiness

$$f(\succsim_1^*, \succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_k^{**}, \succsim_{k+1}^*, \ldots, \succsim_N^*) = f(\succsim_1^*, \succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_k^*, \succsim_{k+1}^*, \ldots, \succsim$$

and on the right hand side 1 and 2 are matched and 3 and k are matched. This implies that if k switches from $\succsim_k^*$ to $\succsim_k^{**}$, 1 and k are not matched in the $1,k$-marginal mechanism at $(\succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_{k+1}^*, \ldots, \succsim_N^*)$. Since either 1 or k must be the dictator in thier marginal mechanism by earlier arguments, it must be 1 and we have

$$k \in g_1(\succsim_2^{**}, \succsim_3^{**}, \succsim_4', \succsim_5^*, \ldots, \succsim_N^*)$$

and since $2, 3, 4$ can't affect 1's option set we get

$$k \in g_1(\succsim_2', \succsim_3', \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_{k+1}^*, \ldots, \succsim_N^*)$$

for all $\succsim_2', \succsim_3', \succsim_4'$. Since k was arbitrary, together with equation 1.10, we have

$$g_1(\succsim_2', \succsim_3', \succsim_4', \succsim_5^*, \ldots, \succsim_{k-1}^*, \succsim_{k+1}^*, \ldots, \succsim_N^*) = N - \{1\} \tag{1.11}$$

for all $\succsim_2', \succsim_3', \succsim_4'$. This, however, gets us back to the first case since 1 is the marginal dictator among $\{1, 2, 3, 4\}$ at $\succsim_5^*, \ldots, \succsim_N^*$ and if $\succsim_3^\dagger \in P^\uparrow(4)$, $\succsim_4^\dagger \in P^\uparrow(3)$, then we must have a marginal dictator among $\{1, 2, 5, \ldots, N\}$, however the only marginal dictator consistent with equation 1.11 is 1. □

1.5.11 Proof of Lemma 4

Nonbossiness is immediate. Then the result follows from the observation that strategy-proofness and nonbossiness are equivlent to group strategy-proofness, recorded in Propostion 1. □

1.5.12 Proof of Theorem 4 (Gibbard–Satterthwaite Theorem)

Let C be the diagonal and $|\mathcal{O}| \geq 3$.

From Proposition 1, it suffices to show that any group strategy-proof mechanism is dictatorial. We will show this in two steps. First, we will show that for some i, j and some profile $\succsim_{-ij} = (\succsim_k)_{k \neq i,j}$ we have $|I^{ij}(\succsim_{-ij})| \geq 3$. From the characterization of two-agent mechanisms, we will see that $f^{ij}_{\succsim_{-ij}}$ is dictatorial. We will then show that this implies the entire mechanism is dictatorial.

1. Suppose by way of contradiction that for all i, j and all $\succsim_{-ij}$ we have $|I^{ij}(\succsim_{-ij}$ $)| < 3$. First, note that if for all i, j and all $\succsim_{-ij}$ we have $|I^{ij}_{\succsim_{-ij}}| = 1$ then f is

single-valued[31] which contradicts the surjectivity of f. Hence there is at least one pair of agents i, j and $\succsim_{-ij}$ such that $|I^{ij}(\succsim_{-ij})| \geq 2$. For simplicity and without loss, let $i = 1$ and $j = 2$. By assumption then $|I^{ij}(\succsim_{-ij})| = 2$ and without loss assume $I^{ij}(\succsim_{-ij}) = \{a, b\}$. Then there must be a local dictator assigned to the incompatible pairs (a, b) and (b, a). This leaves (up to symmetry) two marginal mechanisms ϕ_1 and ϕ_2 where

$$\phi_1(\succsim_1, \succsim_2) = \begin{cases} a & \text{if } a \succ_1 b \\ b & \text{if } a \prec_1 b \end{cases}$$

and

$$\phi_2(\succsim_1, \succsim_2) = \begin{cases} a & \text{if } a \succ_1 b \text{ and } a \succ_2 b \\ b & \text{otherwise} \end{cases}$$

In the first, agent 1 is a dictator. In the second, b is chosen by default and a is only chosen if both agents prefer it to b. Let c be another object in $\mathscr{O}$. If we let $\succsim_2^* \in \mathscr{P}^{\uparrow}[c, a, b]$ then in either case we have $f(\succsim_1, \succsim_2^*, \succsim_{-1,2}) = a$ if $a \succ_1 b$ and $f(\succsim_1, \succsim_2^*, \succsim_{-1,2}) = b$ if $b \succ_1 a$. We then have that a and b are in $I^{1,3}(\succsim_2^*, \succsim_4, \ldots, \succsim_n)$. As before we have two possible mechanisms and in either one, if $\succsim_3^* \in \mathscr{P}^{\uparrow}[c, a, b]$ we have $f(\succsim_1, \succsim_2^*, \succsim_3^*, \succsim_4, \ldots, \succsim_n) = a$ if $a \succ_1 b$ and $f(\succsim_1, \succsim_2^*, \succsim_3^*, \succsim_4, \ldots, \succsim_n) = b$ if $b \succ_1 a$. Continuing in this way, we get a profile of preferences in which all agents prefer c, but c is not chosen. Since any group strategy-proof map is efficient on its image we must either have that $c \notin im(f)$ or f is not group strategy-proof. Either way we have a contradiction.

2. From the characterization of two-agent mechanisms, if $|I^{1,2}(\succsim_{-1,2})| \geq 3$ we have a single dictator in the marginal mechanism $f^{ij}_{\succsim_{-ij}}$. For simplicity let $i = 1, j = 2$ and assume 1 is the dictator. We will show that for any $\succsim'$, $f(\succsim') = \max_{\succsim'_1} I^{1,2}(\succsim_{-1,2})$. Begin with $f(\succsim'_1, \succsim_2, \ldots, \succsim_n)$. The statement holds by assumption. Now since 1 is the marginal dictator, changing $\succsim_2$ to $\succsim'_2$ cannot change the outcome. Hence the statement holds for $f(\succsim'_1, \succsim'_2, \ldots, \succsim_n)$. Now we have that $I^{1,3}(\succsim'_2, \succsim_4, \ldots, \succsim_n)$ contains $I^{1,2}(\succsim_{-1,2})$ as a subset. Hence there either 1 or 3 is a local dictator. Clearly it must be 1. Therefore 3's announcement cannot change the outcome, so we have $f(\succsim'_1, \succsim'_2, \succsim'_3, \succsim_4, \ldots, \succsim_n) = \max_{\succsim'_1} I^{1,2}(\succsim_{-1,2})$. Continuing in this way gives the desired result. The assumption that f is surjective implies that 1 is a dictator. □

1.5.13 Proof of Theorem 5

If $C^{i,j}$ admits more than one equivalence class we may assign a different local dictator to each class as in Theorem 1. We can then extend this mechanism via any GSD-ordering as in Proposition 4. □

[31] To see that $f(\succsim) = f(\succsim')$, change one preference at a time. No single change can alter f, so we get the result.

Chapter 2

Preface

In this chapter, David S. Ahn and I continue to explore constrained allocation mechanisms. We focus on introducing a large class of mechanism which have a number of desirable properties. These mechanisms are defined algorithmically; the mechanism greedily attempts to match agents with their most-preferred alternatives. When conflicts arise, a local priority rule is used to determine which direction is next pursued. This class of mechanisms generalizes many of the known mechanisms used in practice and is applicable to any constraint. In the final chapter, I instead search for stable mechanisms.

Constraint-Traversing Mechanisms

Joseph Root[1] and David S. Ahn[2]

2.1 Introduction

We introduced the notion of a constraint-traversing mechanism in Chapter 1. Here we extend that idea to arbitrary constraints, and provide an analogue to Proposition 2 which guarantees that if the local compromiser assignment follows a set of rules, the resulting mechanism will be group strategy-proof and Pareto efficient.

2.2 Results

Constraint-traversing mechanisms are often pleasant to work with because the notion of group strategy-proofness is strictly stronger than Pareto efficiency.

Proposition 5. *If a constraint-traversing mechanism is group strategy-proof, it is Pareto efficient. However, a constraint-traversing mechanism can be Pareto efficient, but not group strategy-proof.*

Recall that there is no guarantee that an arbitrary local compromiser assignment induces a mechanism. It is possible that the constraint-traversing algorithm will ask an agent to compromise so much that they exhaust all objects. In this case, the algorithm returns $\emptyset$. We will return to this discussion towards the end of this section, for now proceeding with local compromiser assignments for which this does not happen. If α is such that the constraint-traversing algorithm terminates in an allocation for any preference profile, we say α is **implementable**. An implementable α induces a mechanism f^α, in which every preference profile yields the allocation derived from the associated constraint-traversing algorithm. Conversely, a **constraint-traversing mechanism** is

[1]Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: jroot@econ.berkeley.edu

[2]Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: jroot@econ.berkeley.edu

a feasible mechanism $f : \mathscr{P} \to C$ such that there is some local compromiser assignment α which induces it.

An initial difficulty is that the local constraint assignment may not be unique. In Figure 2.1, the three panels correspond to three different two agent local compromiser assignments. Each is implementable, so induces a two agent allocation mechanism. However, all three local compromiser assignments induce the same mechanism. In panels (II) and (III), only 1 or 2 are listed as compromisers at the allocation (a, a), despite the fact that wherever they compromise, the other agent must compromise next. In panel (I), both agents are asked to compromise immediately.

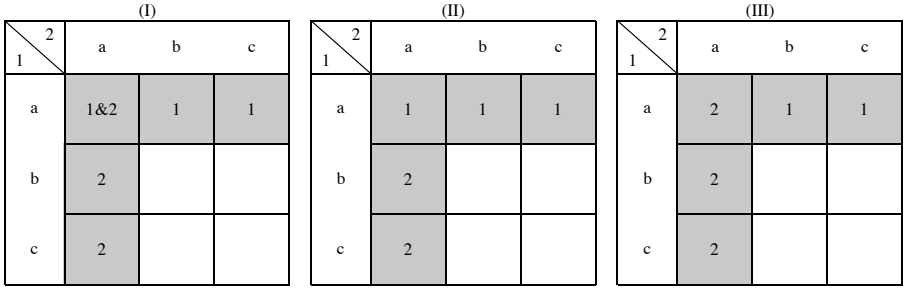


Figure 2.1: Three different local constraint assignments which induce the same mechanism.

Notice, however, that the local compromiser assignment in panel (I) is the pointwise union of the local compromiser assignments in panels (II) and (III). It turns out that this is a general phenomenon when the induced mechanism is group strategy-proof. The pointwise union of all local compromiser assignments which induce a given group strategy-proof mechanism also induces the same mechanism. Furthermore, we show that for any local compromiser assignment, a pointwise nonempty subset, also induces the same mechanism.

Proposition 6. *Let f be a constraint-traversing, group strategy-proof mechanism and let A be the set of local compromiser assignments which induce f. Then A is closed under (pointwise) unions and for any $\alpha \in A$ and α' such that*

$$\emptyset \subsetneq \alpha'(x) \subset \alpha(x) \text{ for all } x \in \bar{C}$$

and $\alpha'(y) = \emptyset$ for all $y \in C$, we have that $\alpha' \in A$.

Definition 5. A local compromiser assignment α is **complete** if for every $x \in \bar{C}$ there is no $i \notin \alpha(x)$ such that $i \in \alpha(y)$ for every y with $x_j = y_j$ for all $j \notin \alpha(x)$.

In words, the local compromiser assignment is complete if there is no agent, not included in $\alpha(x)$ who nevertheless must compromise when the algorithm has reached x [3]. The local compromiser assignment is complete in panel 1 above and is not complete

[3]Since no matter where the local compromisers go, this agent will need to compromise.

in panels 2 and 3. The following proposition shows that for any group strategy-proof mechanism the pointwise union of all local compromiser assignments which induce it is complete.

Proposition 7. *If $f : \mathscr{P} \to C$ is group strategy-proof and constraint-traversing and A is the set of local compromiser assignments which implement f, then $\alpha^* = \cup_{\alpha \in A} \alpha$ is complete.*

Henceforth, when we refer to *the* local compromiser assignment for a given constraint-traversing mechanism we will mean the pointwise union of all local compromiser assignments which induce f.

Another nice feature of group strategy-proof, constraint-traversing mechanisms is that all their marginal mechanisms are also constraint-traversing.

Proposition 8. *Every marginal mechanism of a group strategy-proof constraint-traversing mechanism is constraint-traversing.*

Having done this work, we are now ready to provide sufficient conditions on α for the induced mechanism (provided α is implementable), to be group strategy-proof.

Definition 6. Given a local compromiser assignment α, a **monotone path** is a sequence of allocations

$$z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$$

such that (1) $\{i \,|\, z^l \neq z^{l-1}\} = \{i_l\} \subset \alpha(z^{l-1})$ for all $l = 0, 1, \dots p$ and (2) for all agents i, if $l < m$ and $z_i^l = z_i^m$ then for any $l \leq n \leq m$, we have $z_i^l = z_i^n = z_i^m$

A monotone path is simply a sequence of infeasible allocations (except potentially the last allocation) such that at each step a single agent from the set of local compromisers changes her allocation and such that no agent cycles through objects.

Theorem 6. *If α is implementable and satisfies*

- *[Forward Consistency] For all $x \in \bar{C}$ if $\emptyset \subsetneq A \subsetneq \alpha(x)$ and y is such that $y_j = x_j$ for all $j \notin A$ then $y \in \bar{C}$ and $\alpha(y) \supset \alpha(x) - A$*
- *[Backward Consistency] For all monotone paths,*

$$z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$$

if $j \neq i_1$ is in $\alpha(z^p)$ then for all $x \in \mathcal{O}$, $(x, z^0_{-j}) \in \bar{C}$ and $\alpha(x, z^0_{-j}) \cap \{i_1, \dots, i_{p-1}\}$ is nonempty.

then the induced mechanism f^α is group strategy-proof and Pareto efficient.

Thus far we have simply assumed that the local compromiser assignment is implementable, and therefore induces a mechanism. The following proposition says that it is sufficient to check monotone paths in order to verify that a local compromiser assignment is indeed implementable.

Proposition 9. *A local compromiser assignment α which satisfies forward and backward consistency is implementable if and only if there is no monotone path $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_p} z^p$ in which an agent i compromises $|\mathcal{O}| - 1$ times.*

In the following section we show how this can be used to find mechanisms for a number of constraints.

2.2.1 Examples

No Agent Gets Their Top Choice

Similar to the two-agent case, one might conjecture that the set of constraint-traversing mechanisms when $n > 2$ is in some way related to the set of "local generalized serial dictatorships." These might work as follows: partition the infeasible allocations in such a way that each partition can be assigned a GSD-ordering without conflicting with the other partitions. As in the two-agent case, the top choice of all agents would determine which GSD-ordering is used and the mechanism would yield ex-post the same outcome as in the local GSD-ordering. The following example demonstrates that, at least for some constraints, there are group strategy-proof mechanisms which do not fall into this category.

3:a

1 \ 2	a	b	c
a	1&2	1	1
b	2	3	
c	2		

3:b

1 \ 2	a	b	c
a	1	1	1
b	2		
c			

3:c

1 \ 2	a	b	c
a	2	1	
b	2		
c	2		

Figure 2.2: A non-GSD mechanism

Consider Figure 2.2. The three panels list all possible allocations of three objects $\{a, b, c\}$ to 3 agents. 1's allocation is determined by the row, 2's allocation is determined by the column and 3's allocation is determined by the panel. Grey squares are infeasible and white squares are feasible. For example, the allocation (b, b, a) is infeasible, but the allocation (a, c, c) is feasible. Also listed in Figure 2.2 is a local compromiser assignment which determines a mechanism. Both forward and backward compatibility can be easily checked, giving the following lemma:

Lemma 7. *The mechanism introduced in Figure 2.2 is group strategy-proof and Pareto efficient.*

To see that this is not a local generalized serial dictatorship, consider the preference profile $a \succ_i b \succ_i c$. We start with (a, a, a), move to (b, b, a) and finally to

(b, b, b) which is the outcome. Notice that no agent is getting her top choice, despite the fact that it is possible for all agents to get a[4]. Hence this is inconsistent with any GSD-ordering starting at (a, a, a).

The example is important because it illustrates that constraint-traversing mechanisms are strictly larger than the class of generalized serial dictatorships. Serial dictatorships are sometimes criticized for their lack of fairness in privileging the agents who choose first. In fact, there are constraint-traversing mechanisms that force all agents to compromise.

Variations on the House Allocation Problem

Pycia and Ünver (2017) provide a full characterization of all group strategy-proof and Pareto efficient mechanisms for the house allocation problem. With just three agents and three objects, the house allocation constraint can be visualized as in Figure 2.3. In this section, we will make a slight perturbation to this constraint. With this small perturbation, all existing analyses of the house allocation problem are now inapplicable. However, by traversing the constraint in the way we just described, we can find non-trivial group strategy-proof and Pareto efficient mechanisms for this problem that are not generalized serial dictatorships. This is a "proof of concept" exercise to concretely illustrate how constraint-traversing mechanisms can be constructed for a reasonable-looking problem that would have been otherwise unsolvable. Moreover, the resulting mechanism is of some interest on its own, since it illuminates how tensions between property rights and efficiency are adjudicated by the mechanism in the present of a slightly relaxed constraint.

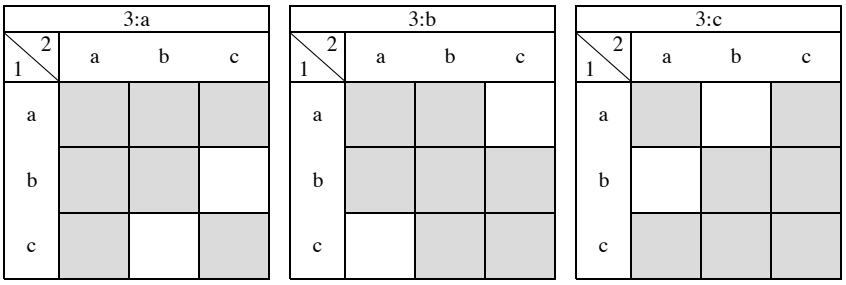


Figure 2.3: Constraint for House Allocation Problem

In Figure 2.4 we list (up to a relabeling of the agents and objects) the set of local compromiser assignments which satisfy both consistency conditions. We drop the labels above to make the figure more compact. These mechanisms have a number of interesting properties. When, for example we list "1 or 2," we mean that the cell can be filled with a '1" or a '2", but not both.

It turns out that this is exactly the set of group strategy-proof and Pareto efficient mechanisms characterized by Pycia and Ünver (2017). There are "heirarchical

[4]The allocations (a, c, c), (c, a, b) and (c, c, a) are all feasible

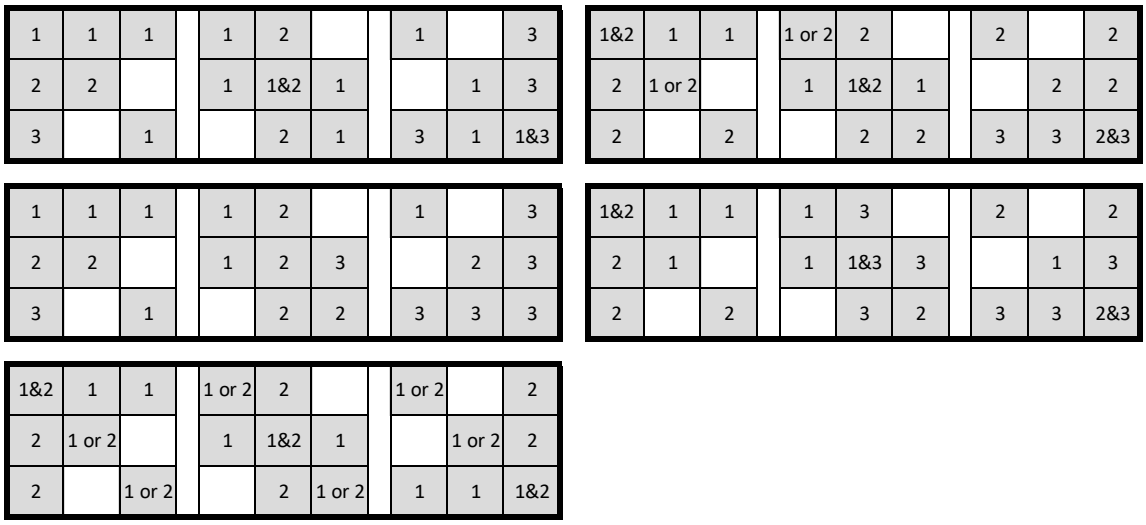


Figure 2.4: All group strategy-proof and Pareto efficient 3-Agent Mechanisms (up to symmetry)

exchange" mechanisms as in Papái (2000) and "broker" mechanisms as in Pycia and Ünver (2017). We show how to find the set of local compromiser assignments in the supplemental appendix.

Suppose, however, that the constraint is as shown in Figure 2.5.

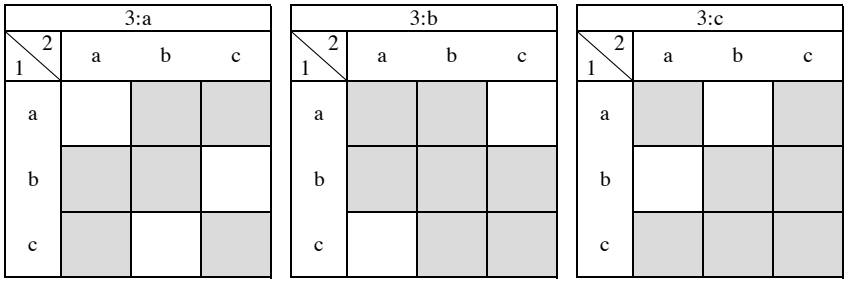


Figure 2.5: A Variation on the House Allocation Constraint

Now, the allocation (a, a, a) is feasible. Otherwise the constraint is exactly the same. Without Theorem 6, one would need to find a way to modify the proof of Pycia and Ünver (2017) to this constraint. In light of this theorem, however, we can simply find the set of local compromiser assignments which satisfy forward and backward consistency. These are listed in Figure 2.6. We demonstrate the process for constructing all of them in the Supplemental Appendix.

These mechanisms demonstrate many of the qualities observed in house allocation problems. In the first mechanism, we have a broker as in Pycia and Ünver (2017). The second mechanism is a mix between serial dictatorship and top trading cycles. The mechanism behaves as though agent 2 owns object b and agent 3 owns object c.

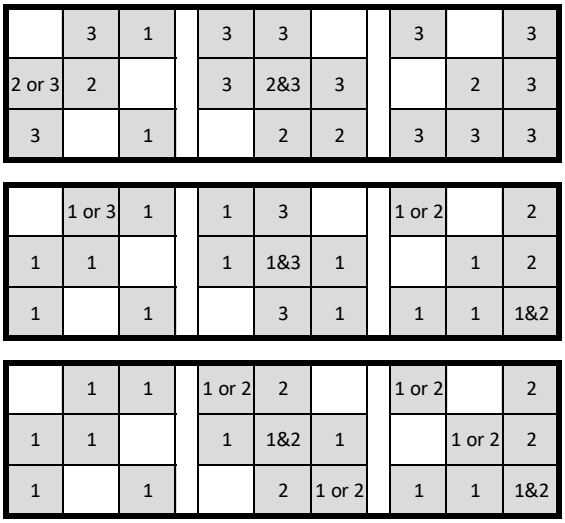


Figure 2.6: The Constraint-Traversing Mechanisms for the Constraint in Figure 2.5

If both agents 2 and 3 top-rank a, then the social allocation is (a, a, a) regardless of 1's preferences. However, 1 can also has some power. If we opt for 3 in the square labeled "1 or 3" and 2 in the square labeled "1 or 2", whenever either 2 or 3 top-ranks the object she owns and there is a conflict between the other two agents over a, in this case 1 forces the other agent to compromise. This demonstrates that simply following the consistency conditions to construct constraint-traversing algorithms can yield mechanisms with interesting properties.

2.3 Appendix

2.3.1 Proof of Proposition 5

By construction, a constraint traversing mechanism's image is C. By Lemma 1, we thus have that if f group strategy-proof, it is also Pareto efficient. The following mechanism is easily seen to be Pareto efficient, but letting $a \succ_3 b \succ_3 c$ the marginal mechanism for agents 1 and 2 is not group strategy-proof[5]. □

3:a

1 \ 2	a	b	c
a	1	3	3
b	2		
c			

3:b

1 \ 2	a	b	c
a			
b			
c			

3:c

1 \ 2	a	b	c
a			
b			
c			

2.3.2 Proof of Proposition 6

The proofs of these two statements are nearly identical. Let α and α' induce f which is group strategy-proof. Let $\succsim$ be an arbitrary preference profile and set $\succsim^0=\succsim$. Iteratively define the sequence $\succsim^0, \succsim^1, \ldots, \succsim^N$ so long as $\tau(\succsim^n) \notin C$ by $\succsim_i^{n+1}=\succsim_i^n$ for all $i \notin \alpha \cup \alpha'(x)$ and $\succsim_j^{n+1}$ is identical to $\succsim_j^n$ except $\tau(\succsim_j^n)$, is sent to the bottom for all $j \in \alpha \cup \alpha'(x)$. At each step, we have $f_j(\succsim^n) \neq \tau(\succsim_j^n)$ for all $j \in \alpha \cup \alpha'(\tau(\succsim^n))$ so Maskin monotonicity implies that $f(\succsim^n) = f(\succsim^{n+1})$. However the sequence $\tau(\succsim^n)$ is precisely the set of allocations achieved in the constraint traversing algorithm under $\alpha \cup \alpha'$. The algorithm ends at the first feasible assignment, and since $f(\succsim^n)$ is unchanged throughout the process, we get that $f(\succsim^N) = f^{\alpha \cup \alpha'}(\succsim)$ which gives the result. To prove the second claim, set $\succsim^0=\succsim$ as before and again iteratively define the sequence $\succsim^0, \succsim^1, \ldots, \succsim^N$ so long as $\tau(\succsim^n) \notin C$ by $\succsim_i^{n+1}=\succsim_i^n$ for all $i \notin \alpha'(x)$ and $\succsim_j^{n+1}$ is identical to $\succsim_j^n$ except $\tau(\succsim_j^n)$, is sent to the bottom for all $j \in \alpha'(\tau(\succsim^n))$. Maskin monotonicity implies that f^α is unchanged along the sequence and again the sequence of $\tau(\succsim^n)$ follows the allocations in the steps of the constraint-traversing mechanism. □

2.3.3 Proof of Proposition 7

Let $x \in \bar{C}$ and i be an agent such that whenever x is top ranked (i.e x_i is top-ranked for each i), must always compromise. That is, if $\succsim$ is any profile such that $x_j = \max_{\succsim_j} \mathcal{O}$ for all j then $f_i^\alpha(\succsim) \neq x_i$. Let α' be identical to α except that $\alpha'(x) = \{i\}$. We claim that for every preference profile the constraint traversing algorithm using α and

[5]The cells filled with a number are the infeasible allocations. For example, (a, a, a) and (b, a, a) are infeasible, but (b, b, a) is feasible.

α' yield the same allocation, so that α' is implementable and $f^\alpha = f^{\alpha'}$. Of course, the constraint traversing algorithm for α and α' can only yield different outcomes for preference profiles $\succsim$ in which the constraint-traversing algorithm lands at x at some point. Any such preference profile must satisfy $x_j \succsim_j f_j^\alpha(\succsim)$ for all j. Given such a profile $\succsim$, define $\succsim'$ so that agent $j \neq i$ has the preference $\succsim'_j$ defined by

$$x_j \succsim'_j LC_{\succsim_j}(x_j) \succsim'_j UC_{\succsim_j}(x_j)$$

where the ranking within groups is identical to $\succsim_j$ and i has the preference $\succsim_i$ defined by

$$LC_{\succsim_i}(x_i) \succsim'_i UC_{\succsim_i}(x_i) \succsim'_i x_i$$

but again the ranking within the three sets is determined by $\succsim_i$. Maskin monotonicity implies that $f^\alpha(\succsim) = f^\alpha(\succsim')$ since we have only reduced the upper contour sets of the $f_j^\alpha(\succsim)$. And since x_i is i's last choice, the constraint-traversing algorithm for $\succsim'$ never lands on x so we have $f^{\alpha'}(\succsim') = f^\alpha(\succsim')$. Finally, the algorithm under α' at $\succsim$ eventually lands on x at which point i compromises. After that, the algorithm operates identically to the algorithm under α at $\succsim'$, and therefore yields the same outcome. □

2.3.4 Proof of Proposition 8

Let f be a constraint-traversing mechanism for the local constraint assignment α. Pick any proper subset M of N and a preference profile $\succsim_{M^c}$ for the other agents. We must show that $f^M_{\underset{\sim}{C}_M}$ is constraint-traversing on $I^M(\succsim_{M^c})$ for some local compromiser assignment. First, let $x^M \in \mathcal{O}^M - I^M(\succsim_{M^c})$ so that x^M is a suballocation for the agents in M which is unacheivable under f when the agents in M^c announce the preference profile $\succsim_{M^c}$. Let $\succsim_M$ be a preference profile such that $\tau(\succsim_M) = x^M$. Now $\tau(\succsim_M, \succsim_{M^c})$ is infeasible because otherwise by Pareto efficiency we would have $f(\succsim_M, \succsim_{M^c}) = \tau(\succsim_M, \succsim_{M^c})$ and hence $x^M \in I^M(\succsim_{M^c})$. Let x^* be the first allocation in the constraint-traversing algorithm at $(\succsim_M, \succsim_{M^c})$ under α in which $\alpha(x^*) \cap M$ is nonempty. Such a point is guaranteed again because $x \in \mathcal{O}^M - I^M(\succsim_{M^c})$. Define $\alpha^*(x) = \alpha(x^*) \cap M$. This choice is independent of the choice of $\succsim'_M$ with the property that $\tau(\succsim_M) = x$, since the constraint-traversing algorithm under α, until reaching x^*, only depends on the top choices of the agents in M. Thus we may define α^* likewise on the rest of $\mathcal{O}^M - I^M$ in a well-defined way. It remains to show that α^* implementable and that the induced algorithm agrees with $f^M_{\succsim_{M^c}}$.

To see this, pick an arbitrary $\succsim_M$. If $\tau(\succsim_M) \in I^M_{\succsim_{M^c}}$, then the constraint traversing algorithm under α^* gives the suballocation $\tau(\succsim_M)$, which agrees with $f(\succsim_M, \succsim_{M^c})$ by group strategy-proofness. Otherwise, $\tau(\succsim_M) \notin I^M_{\succsim_{M^c}}$. Now by definition we have that the agent(s) in $\alpha^*(\tau(\succsim_M))$ cannot get their top choice under f at the profile $(\succsim_M, \succsim_{M^c})$. We can therefore modify $\succsim_M$ to $\succsim^2_M$ by having each agent in $\alpha^*(\tau(\succsim_M))$ move their top choice to the bottom. Now Maskin monotonicity implies that $f(\succsim^2_M, \succsim_{M^c}) = f(\succsim_M, \succsim_{M^c})$. If $\tau(\succsim^2_M) \in I^M(\succsim_{M^c})$, we stop. Otherwise, we repeat the process. Continuing in this way, we get a sequence of profiles $\succsim^1_M, \succsim^2_M, \ldots, \succsim^n_M$

with $f(\succsim_M^k, \succsim_{M^c}) = f(\succsim_M^l, \succsim_{M^c})$ for all l, k. Furthermore, the sequence $\tau(\succsim^i)$ follows the allocations in the constraint-traversing mechanism under α^* exactly. Thus $f_M(\succsim_M^n, \succsim_{M^c}) = \tau(\succsim_M^n)$ and $\tau(\succsim_M^n)$ is the outcome of the constraint-traversing algorithm under α^*. □

2.3.5 Proof of Theorem 6

We show this in four steps. First, we show that if α satisfies forward consistency, then any α' such that for all $x \in \bar{C}$, $\emptyset \subsetneq \alpha'(x) \subset \alpha(x)$ also implements f^α. Next, we show that this, along with forward consistency imply that the marginal mechanisms holding a single agents' preferences fixed for f^α are all constraint-traversing. Third, we show that these constraint traversing mechanisms also satisfy forward and backward consistency. Finally, we establish the result by showing that forward and backward consistency imply the group strategy-proofness of the two-agent marginal mechanisms.

To see that the set of local compromiser assignments which induce f is closed under (nonempty) pointwise subsets, let $x \in \bar{C}$ such that $\alpha(x)$ is multi-valued with $i \in \alpha(x)$. Define $\alpha'(y) = \alpha(y)$ for all $y \neq x$ and $\alpha'(x) = \alpha(x) - \{i\}$. Of course for any preference profile such that the constraint-traversing algorithm under α never lands on x will yield the same result under α'. Let $\succsim$ be a preference profile such that the constraint-traversing algorithm under α eventually lands on x. Then the sequence of allocations achieved in both algorithms is identical until they both land on x. Let $z^0, z^1, \ldots z^p$ be the sequence of allocations after x in the constraint-traversing algorithm under α and $w^0 \to w^1 \to \cdots w^q$ be the same sequence for α'. Now $x = z^0 = w^0$ and w^1 Pareto-dominates z^1 (because fewer agents had to compromise). However, by forward consistency z^1 (weakly) Pareto-dominates w^2. Again applying forward consistency we have that w^2 Pareto-dominates z^2. Continuing this logic forward we have

$$w^1 \geq_{PD} z^1 \geq_{PD} w^2 \geq_{PD} z^2 \geq_{PD} w^3 \cdots$$

and whichever sequence stops first, the other one has to stop at the same time since otherwise we would get a contradiction to forward consistency. Thus we have that the constraint-traversing algorithm under α and α' result in the same outcome at $\succsim$. Hence $f^\alpha = f^{\alpha'}$. Of course, for any α'' such that $\emptyset \subsetneq \alpha''(x) \subset \alpha(x)$ on $\bar{C}$ we can iteratively remove one agent at a time, to get that α'' implements α.

Next, Let h be the marginal mechanism holding agent k at $\succsim_k$. We want to show that h is constraint-traversing. To do so, for every $x \in \bar{C}$ define

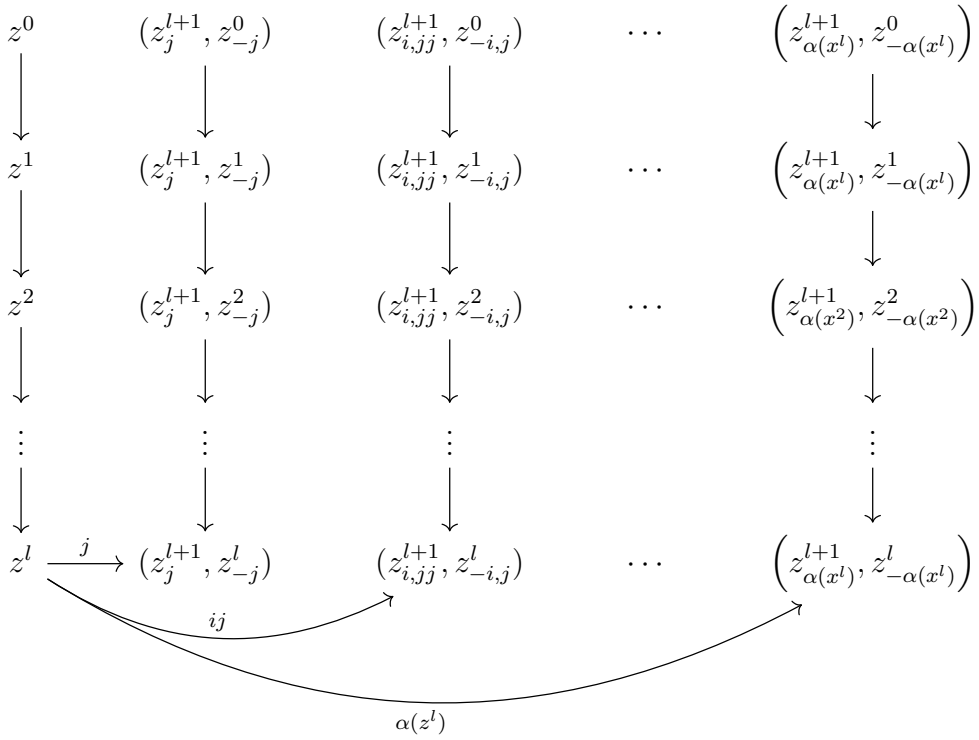
$$\alpha'(x) \begin{cases} \alpha(x) & \text{if } k \notin \alpha(x) \\ k & \text{if } k \in \alpha(x) \end{cases}$$

by the result above, α' implements α. For any suballocation z of the agents other than k, define $\alpha^*(z) = \alpha'(y)$ where y is the first allocation in the sequence $(\tau_n(\succsim_k), z)$ such that $\alpha'(y) \neq \{k\}$. We want to show that α^* implements h. To do so, we will induct on the number of steps required in the constraint-traversing algorithm under α'. If $\succsim$ is a

preference profile such that the constraint-traversing algorithm under α' takes just one step, then $\tau(\succsim_{-k}, \succsim_k)$ is feasible, so $h(\succsim) = \tau(\succsim)$ which is the outcome of $\succsim$ under α^*. Now assume that the statement holds for all preference profiles which take less than or equal to n steps under α'. Let $\succsim_{-k}$ be a preference profile such that the outcome of $(\succsim, \succsim_k)$ takes $n+1$ steps under α'. Let $z^0 \to \cdots z^n$ be these steps. If $\alpha'(z^0) \neq \{k\}$ then $\alpha^*(z^0) = \alpha'(z^0)$. Let $\succsim'_{-k}$ be the profile in which each agent from $\alpha'(z^0)$ puts their top choice to the bottom of their list, without changing anything else. Then, by design, the sequence of allocations in the constraint-traversing algorithm under α' is $z^1 \to z^n$ which takes only n steps. Thus by the induction assumption, $f(\succsim'_{-k}, \succsim_k)$ is the same allocation as we get from running the constraint-traversing algorithm under α^* at $\succsim_{-k}$. However, we also have that $f(\succsim'_{-k}, \succsim_k) = f(\succsim_{-k}, \succsim_k)$ by construction and the outcome of α^* under $\succsim_{-k}$ is the same as under $\succsim'_{-k}$ since the latter simply skips the first step. This gives the desired result. Suppose now that $\alpha'(z^1) = \{k\}$. If the same holds for the entire sequence, we again get $f(\succsim_{-k}, \succsim_k) = \tau_1(\succsim_{-k}, \succsim_k)$ hence $h(\succsim_{-k}) = \tau(\succsim_{-k})$ which is the same as we get from α^*. Finally suppose that $\alpha'(z^0) = \{k\}$ but that there is a $l \geq 1$ such that $\alpha'(z^l) \neq \{k\}$. Assume that l is the minimum index such that this holds. Now for all $m \leq l$ we have a monotone sequence $z^m \xrightarrow{k} z^l$. Let $j \in \alpha'(z^l)$. By backward consistency for all $x \in \mathcal{O}$ we have $(x, z^l_{-j}) \in \bar{C}$ and $k \in \alpha(x, z^l_{-j})$. In particular, we have $k \in \alpha(z^{l+1}_j, z^l_{-j})$. But then we have a monotone path $(z^{l+1}_j, z^l_{-j}) \xrightarrow{k} (z^{l+1}_j, z^{l+1}_k, z^l_{-j,k})$. Furthermore from forward consistency, $\alpha(x^l) - \{j\} \subset \alpha(z^{l+1}_j, z^{l+1}_k, z^l_{-j,k})$. Hence we can continue this way, replacing the object for each agent in $\alpha(z^l)$ by the object they receive in z^{l+1}.This process is illustrated in the following diagram:

the idea is to remove the first step of the algorithm under α^*, which is not the first step of the algorithm under α'. However, the first steps under α' are just k compromising. Thus, as shown in the figure, we can use backward consistency to show that the we can one-at-a-time move the agents in $\alpha^*(\tau_1(\succsim_{-k}))$ to their second best choice. Having done this, if we let $\succsim'_{-k}$ be the preference profile in which all agents in $\alpha^*(\tau_1(\succsim_{-k}))$ put their top choice to the bottom, without changing anything else. Then by design the outcome at $\succsim'_{-k}$ under α^* is the same as $\succsim_{-k}$ since again we just skip the first step. But we also from the argument above that $f^\alpha(\succsim_{-k}, \succsim_k) = f^\alpha(\succsim'_{-k}, \succsim_k)$ and, for the agents other than k, the latter is the same as the outcome of α^* at $\succsim_{-k}$. This gives the desired result.

For step 3, we need to show that α^* defined above satisfies forward and backward consistency. We will start with the easier of the two: forward consistency. Suppose that $\alpha^*(x)$ is multi-valued and $\emptyset \subsetneq A \subsetneq \alpha(x)$ and y is such that $y_j = x_j$ if $j \notin A$. Let q_k be k's object in the first allocation along the sequence $(\tau_n(\succsim_k), x)$ where $\alpha'((\tau_n(\succsim_k), x))$ is not k, i.e we have $\alpha'(q_k, x) = \alpha^*(x)$. By forward consistency of α and since $\alpha(q_k, x) = \alpha'(q, x)$, we have that $\alpha(q_k, y) \supset \alpha(q_k, x) - A$. Now we need to show that the same holds for $\alpha^*(y)$. However, this follows from a similar process to the last step. If $k \notin \alpha(q_k, y)$, then we can repeat exactly the process before to show that the first allocation in the sequence $(\tau_n(\succsim_k), y)$ is exactly (q_k, y). Then from the definition of α^* we get the result. Otherwise, $k \in \alpha(q_k, y) \supset \alpha(q_k, x) - A$ so that $\alpha'(q_k, y) = \{k\}$. But in this case, we do



the same and continue forward and apply forward compatibility once more to find the result. Next we need to show that α^* satisfies forward consistency. First we will need a definition. Given a monotone sequence $z^{(n)} = z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_p} z^p$ we let $\succsim_{z^{(n)}}$ be the preference profile in which all agents put z^0 as their top choice. i_1 puts $z_{i_1}^1$ as her next best choice and so on. We will also need the following proposition:

Proposition 10. *If η satisfies forward and backward consistency and $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_p} z^p$ is a monotone sequence for η and $i_p \neq i_l$ with $l < p$ then there is a monotone sequence $w^{(n)}$ starting at $(z_{-i_p}^0, z_{i_p}^p)$, ending at z^p and such that $\succsim_{w^{(n)}}$ is the same as $\succsim_{z^{(n)}}$ except that agent i_p puts her top option to the bottom*

Proof. We will proceed by induction on the length of the monotone sequence. If the sequence has just one step, then the result is trivial since the new monotone sequence is just a single element (namely z^1). Suppose that for $m \leq n$ if the sequence has m steps, the proposition holds. Let $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_{n+1}} z^{n+1}$ be a monotone sequence such that i_{n+1} hasn't compromised before. By backward consistency $\eta(z_{-i_{n+1}}^0, z_{-i_{n+1}}^{n+1})$ intersects $\{i_1, \ldots, i_n\}$. Suppose specifically that $i_l \in \eta(z_{-i_{n+1}}^0, z_{-i_{n+1}}^{n+1})$. By the induction hypothesis, there is a monotone sequence $w^{(n)}$ of length $l-1$ which starts at $(z_{-i_l}^0, z_{i_l}^l)$ and ends at z^l. We can continue this monotone sequence so that it follows $z(n)$ after landing on z^l. This gives a monotone sequence of length n so by the induction hypothesis again we get a monotone sequence from $(z_{-i_l,i_{n+1}}^0, z_{i_l}^l, z_{-i_{n+1}}^{n+1}$ to z^{n+1}. However, this can now just be the second step of a new monotone sequence that starts at $(z_{-i_{n+1}}^0, z_{-i_{n+1}}^{n+1})$.

This gives the desired result. □

Now, we are ready to show that α^* satisfies forward consistency. Given a monotone sequence $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_p} z^p$ under α^*, we can extend this sequence to the sequence of allocations traversed under $alpha'$ at the profile $(\succsim_{z^{(n)}}, \succsim_k)$ to get w^n. Now this has only (potentially) added steps where k compromises. Assume that w^n starts with z^0 since otherwise it starts with a number of allocations where k compromises until we land at z^0 and it will make no difference in the following analysis. Likewise, we may assume that the last step of $w^{(n)}$ is z^p. By backward consistency, we have that $(z^0_{-i_p}, x)$ is in $\bar{C}$ and that there is a local compromiser at this allocation from the set $\{i_1, \ldots, i_n\} \cup \{k\}$. If the local compromiser is not k, then a simple argument similar to the one shown in the diagram above gives that $\alpha^*(z^0_{-i_p}, x)$ also intersects $\{i_1, \ldots, i_n\}$. Otherwise, it is k at which point we apply the proposition above to find that we eventually land on z^l, but this means that at some point we landed at an allocation in which an agent other than k must have compromised.

Finally, we may prove the result. To do so, we simply take margins until we get to every 2-agent marginal mechanism. By the results above this will satisfy forward and backward consistency. However, this immediately implies that they are local dictatorships, which gives the result. □

2.3.6 Proof of Proposition 9

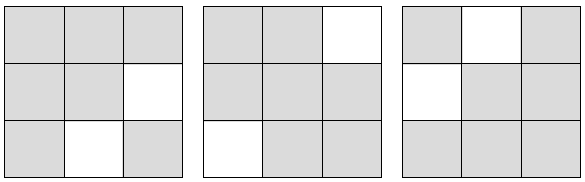
Suppose that α is implementable and satisfies forward and backward consistency. Then by Theorem 6, the induced mechanism is group strategy-proof. Suppose by way of contradiction, there is a $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_p} z^p$ in which an agent k compromises $|\mathcal{O}| - 1$ times. For each i, let $\succsim_i$ top-rank z^0_i, put z^1_i next (if it's different) and so on. Then by Proposition 6 we can let $\alpha'(z^0) = i_1$, $\alpha'(z^1) = i_2$ and so on and the induced mechanism under α' is the same as under α. However the constraint-traversing algorithm for α' under $\succsim$ runs out of k's possible allocations. Hence α' is not implementable, α is not implementable, a contradiction.

Now suppose that α satisfies forward and backward consistency and here is no monotone path $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \cdots \xrightarrow{i_p} z^p$ in which an agent i compromises $|\mathcal{O}| - 1$ times. Suppose by way of contradiction, that α is not implementable. Then there is an agent k and a preference profile $\succsim$ such that the constraint-traversing algorithm runs out of objects for k. We will construct a monotone path with the property we ruled out. Let $w^0 \to w^1 \cdots \to w^q$ be the sequence of allocations. At each step, a number of agents compromise. However, by forward consistency, we can extend this to a monotone path (in which a single agent compromises at each step) with the same properties. This gives the desired violation. □

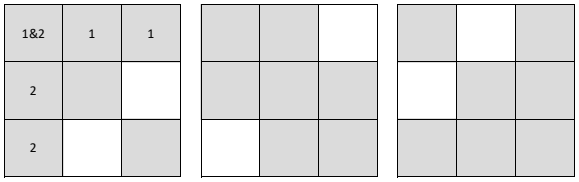
2.4 Supplemental Appendix

2.4.1 3-Agent House Allocation

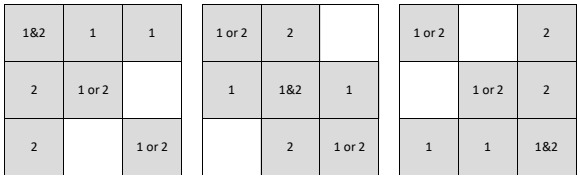
In what follows, we show how to construct the set of 3 agent house allocation mechanisms which are constraint-traversing. We show how to identify the set of local compromiser assignments which satisfy both forward and backward consistency. We will heavily exploit the symmetry of the constraint to reduce the workload. The three objects are labeled a,b and c. Each row corresponds to a different allocation for agent 1, with the top row being a, the second row b and the third row c. Each column corresponds to 2's allocation with the first column a and so on. The three panels correspond to 3's allocation with the first panel being object a and so on.



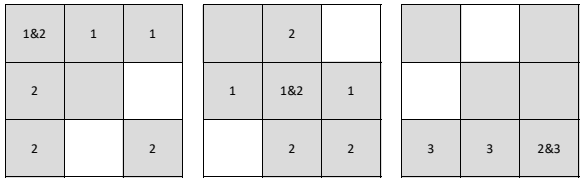
(1) This is the constraint. The grey squares are infeasible and the white squares are feasible.



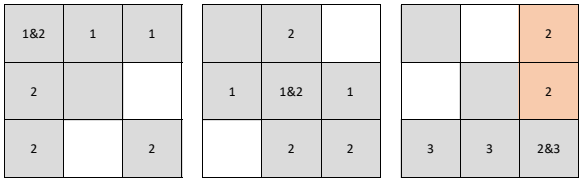
(2) At the allocation (a,a,a), at least two agents will need to compromise. We don't need to list two agents at this square as we will see later. For now, however, we will and without loss, we'll choose agents 1 and 2. The other squares are filled out according to forward consistency.



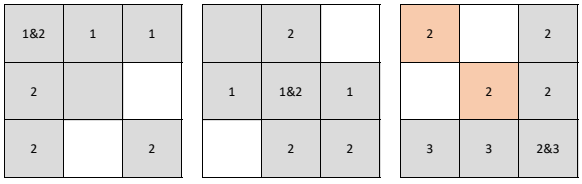
(3) The allocations (b,b,b) and (c,c,c) also need two compromisers. We'll start by assuming all three are the same: 1 and 2 are asked to compromise. Then the rest of the squares are filled out via forward consistency. This satisfies both forward and backward consistency, so gives a GSP, PE mechanism.



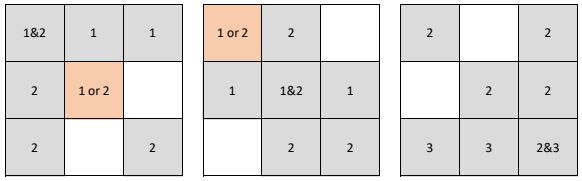
(4) Here, instead we have (2,3) compromise at (c,c,c).



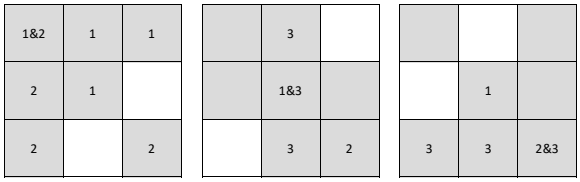
(5) The highlighted squares can't be 1 by backward consistency, since in either case we'd have a monotone path from this square to (c,c,c) which has 2 as a label. This would violate backward consistency because in both cases we can move in 2's direction and encounter a feasible allocation. If they were labeled 3, we'd get a 3-step monotone path to a 2 in either case.



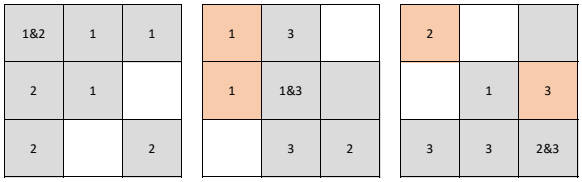
(6) The highlighted squares can't be labeled 3 since we get an immediate violation of backward consistency. If either were labeled 1, we would also get a violation of backward consistency for the monotone path (a,c,c) --> (a,a,c) -- > (c,a,c) and the monotone path (b,c,c) --> (b,b,c) -- > (c,b,c) respectively.



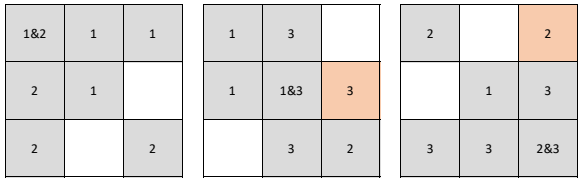
(7) Finally, the last two unfilled squares can't be labeled with a 3 since both lead to immediate violations of backward consistency. However, both squares can be labeled 1 or 2 independently without any violation of the consistency conditions. This gives 4 mechanisms.



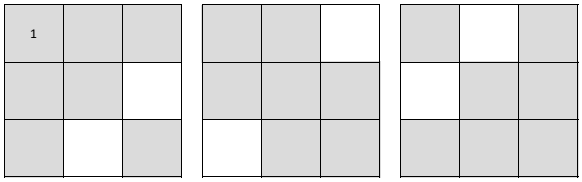
(8) Finally, we can have all three diagonal squares with different pairs of compromisers.



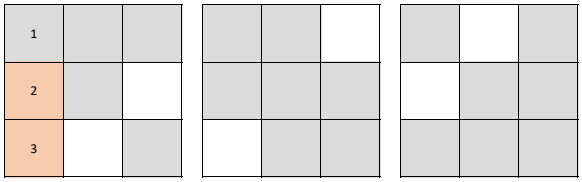
(9) If we label either square in the second panel 3 or 2 we get a 3 step monotone path to a "1" which can't satisfy backward consistency. The square (a,a,c) can't be labeled 3 because we get an immediate violation of backward consistency. It can't be labeled 1 because we get a 3-step monotone path to a "2." The final square is similar.



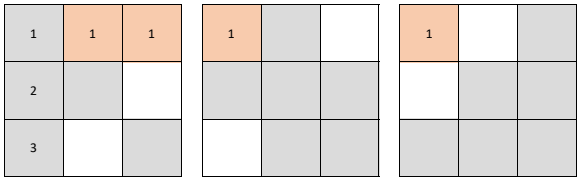
(10) The square (b,c,b) clearly can't have a 2 and it can't have a 1 because of the monotone path (b,c,c) -- > (b,c,b) -- > (c,c,b). The top right square can't have a 1 or a 3 because both lead to immediate violations of backward consistency. This gives a mechanism which is GSP and PE by the constraint-traversing theorem.



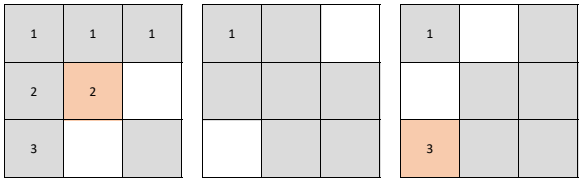
(11) Before we assumed that two agents were listed as compromisers at each of the squares (a,a,a), (b,b,b) and (c,c,c). Instead we assume here that at least one has only one agent listed. We will deduce the rest of the mechanism given this assumption.



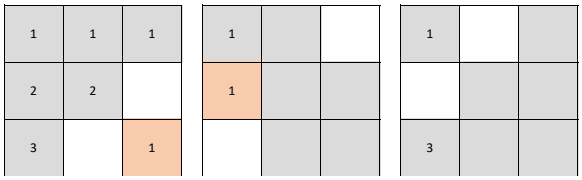
(12) By completeness the two highlighted squares can't have the same label. By symmetry we fill it out without loss as above.



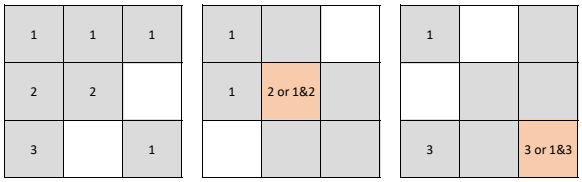
(13) The highlighted squares follow from backward consistency.



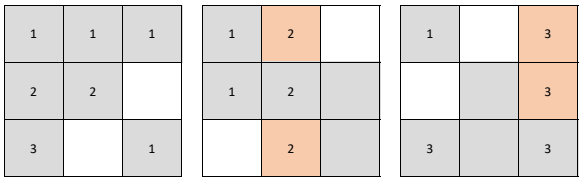
(14) The highlighted square in the left panel can't be labeled "3" because of the "1" above it. It can't be labeled 1 because backward consistency would require (a,a,a) also have 2 as a compromiser. The other square is symmetric.



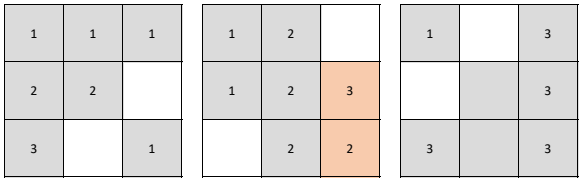
(15) The highlighted square in the second panel can't be labeled "3" because it would lead to 3 step monotone path starting at (a,a,b) and ending at (b,a,a) which cant satisfy backward consistency. The other square is symmetric.



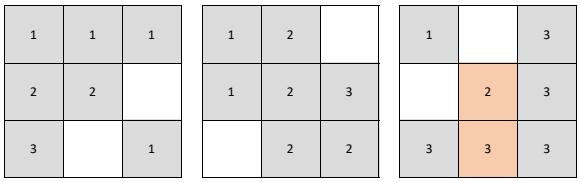
(16) Now we have four potential cases. The squares (b,b,b) and (c,c,c) can be labeled as shown above. (b,b,b) can't have a 3 because backward consistency would require a 3 in the squares (b,x,b) which can't be efficient, since 1 is already listed at (b,a,b).



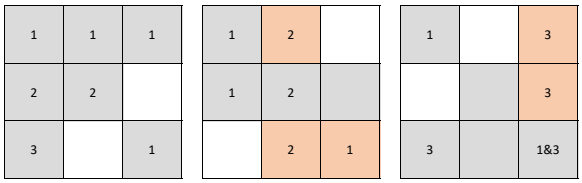
(17) We'll start with a "2" at (b,b,b) and "3" at (c,c,c). Backward consistency leads to the immediate labels in the highlighted square.



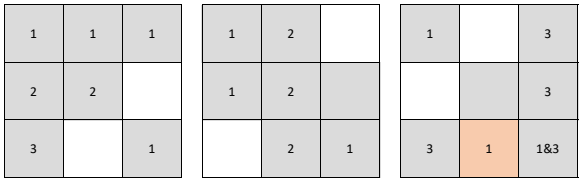
(18) By completeness, the two highlighted squares have to be as shown. For example, if (c,c,b) were labeled "1" then we would have to label (c,c,c) 1&3 by completeness.



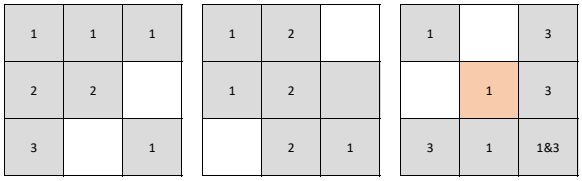
(19) The square (c,b,c) has to be labeled "3" by backward consistency. Then the other square can't be labeled "1" because we would get a 3 step violation from the monotone path starting there and ending at (c,b,b). Since this local compromiser assignment satisfies the consistency conditions, the associated mechanism is GSP and PE



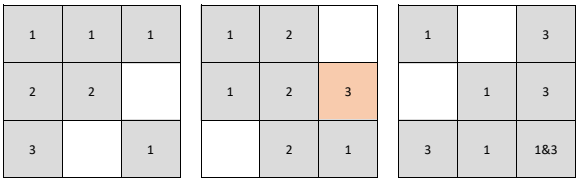
(20) Now we label (b,b,b) with a "2" and (c,c,c) with a "1&3." Then the highlighted squares follow from backward and forward consistency.



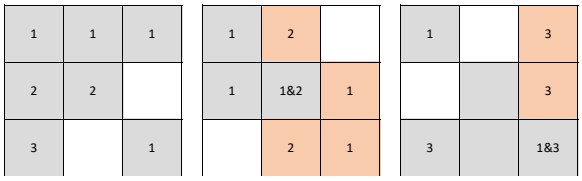
(21) The highlighted square can't be 3 or 2 because either give a violation of backward consistency. The former because of the monotone sequence (c,b,c) --> (c,b,b) -- > (c,c,b).



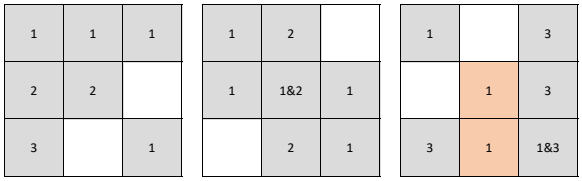
(22) The highlighted square can't have a 1 or a 2 because both lead to immediate violations of backward consistency.



(23) By completeness the final square has to have a 3. However this leads to a violation of backward consistency by the monotone path (c,b,b) --> (c,c,b) -- > (b,c,b). So, this choice doesn't work.



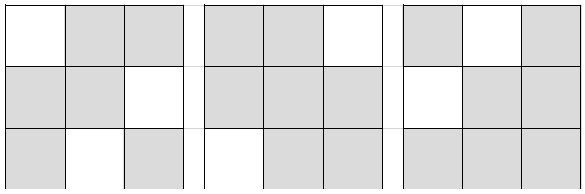
(24) Finally we have the case where the allocations (b,b,b) and (c,c,c) both have two local compromisers listed. The highlighted squares come from forward consistency.



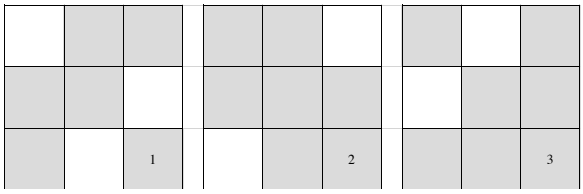
(25) The final squares have to be filled out as follows. If either has a "3" we get a 3-step monotone path which can't satisfy backward consistency. The same happens if either is labeled "2." This labeling satisfies both consistency conditions. Hence, we get a GSP, PE mechanism.

2.4.2 A Variation on the 3-agent House Allocation Constraint

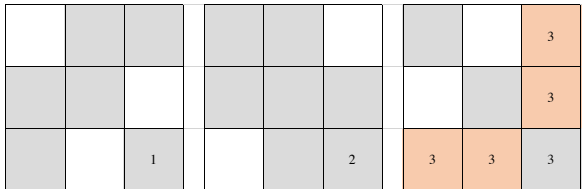
In what follows, we show how to construct the set of constraint-traversing mechanisms for the constraint shown below. This is a variation on the house allocation constraint. We show how to identify the set of local compromiser assignments which satisfy both forward and backward consistency. We will heavily exploit the symmetry of the constraint to reduce the workload. The three objects are labeled a,b and c. Each row corresponds to a different allocation for agent 1, with the top row being a, the second row b and the third row c. Each column corresponds to 2's allocation with the first column a and so on. The three panels correspond to 3's allocation with the first panel being object a and so on. This constraint differs from the house allocation because the allocation (a,a,a) is feasible.



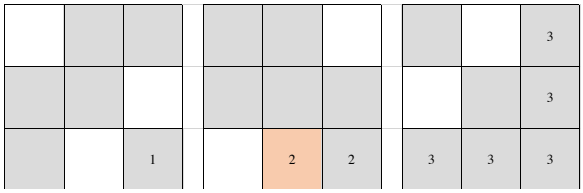
(1) This is the constraint. The difference between this and the house allocation constraint is that the allocation (a,a,a) is feasible. As in the house allocation constraint, we will start with the allocations (b,b,b) and (c,c,c).



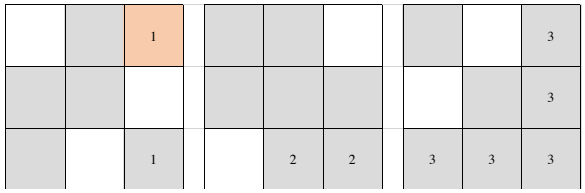
(2) In this case we will start by assuming that there is a single compromiser at(c,c,c). We will study later the implication of assuming more than a single compromiser here. By completeness, if (c,c,a) and (c,c,b) both had the label 1 or 2, the allocation (c,c,c) would have to list more than one agent. By symmetry we'll just choose this arrangement.



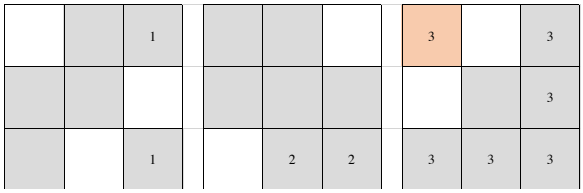
(3) The highlighted squares are implied by backward consistency



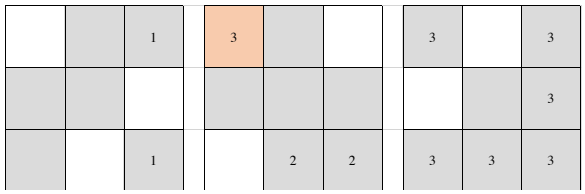
(4) The highlighted square needs to be labeled 2 or 3, but if 3, backward consistency would imply that 2 compromises at (c,c,c).



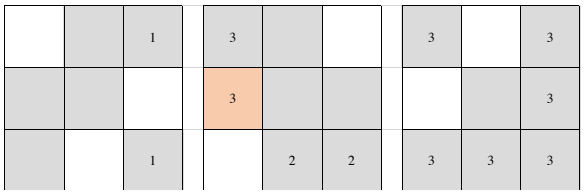
(5) The highlighted square can't be labeled 2 or 3 because both give immediate violatins of backward consistency.



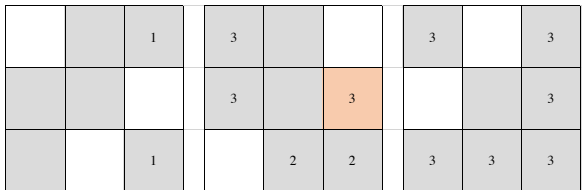
(6) The highlighted square can't have a 1 or a 2 since backward consistency would imply that (a,a,a) should have the same label in both cases.



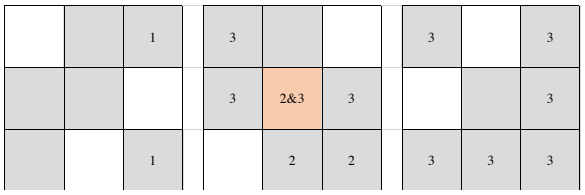
(7) Given the 3 at (a,a,c) the highlighted square cannot have a 1 or a 2.



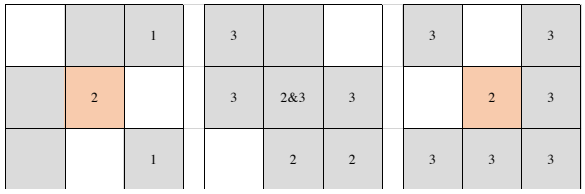
(8) The highlighted square can't have a 2 since eventually another agent will have to compromise leading to a violation of backward consistency. It can't have a 1 because this would lead to an immediate violation of backward consistency.



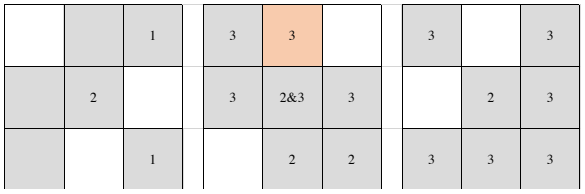
(9) The highlighted square has to be labeled 1 or 3. 1 requires that (b,a,b) be labeled 1 by backward consistency, but it is already labeled 3, and having both doesnt cant satisfy forward consistency.



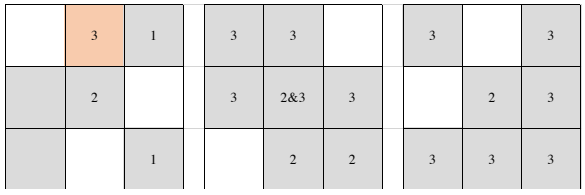
(10) This square can't have a 1 because of the 3's around it. If it has a 2 by completeness it also has a 3. If it only had a 3 then by backward consistency we would have to have (c,b,b) labeled 3, which it is not.



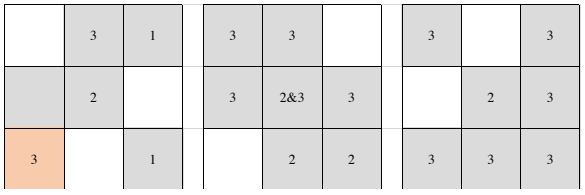
(11) Forward consistency implies the highlighted squares are labeled 2.



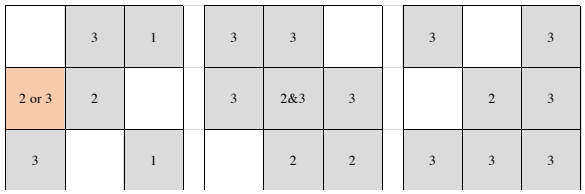
(12) The highlighted square can't have a 1 or a 2 since both lead to immediate violations of backward consistency.



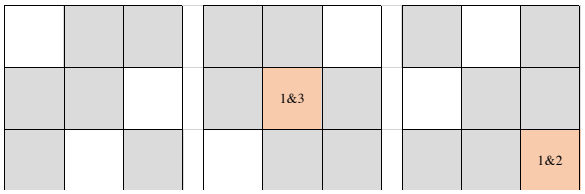
(13) The highlighted square can't have a 1 or a 2 since both lead to immediate violations of backward consistency.



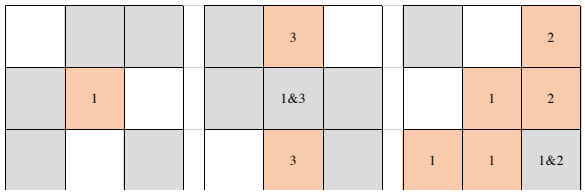
(14) The highlighted square can't have a 1 or a 2 since both lead to immediate violations of backward consistency.



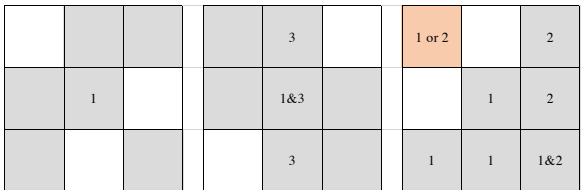
(15) The final square can have a 2 or a 3, but not a 1. Both choices lead to local compromiser assignments which satisfy both consistency conditions.



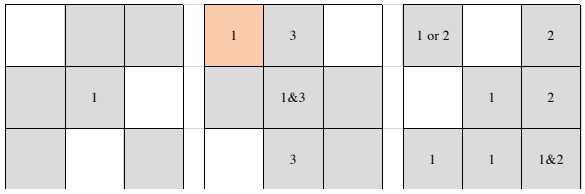
(16) We started before with a single compromiser at (c,c,c) and deduced the entire mechanism. Instead here we assume that both (b,b,b) and (c,c,c) have two compromisers, but in this case the pairs are different.



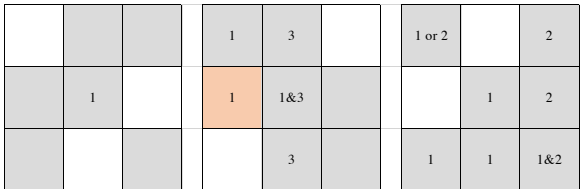
(17) The implications of forward consistency.



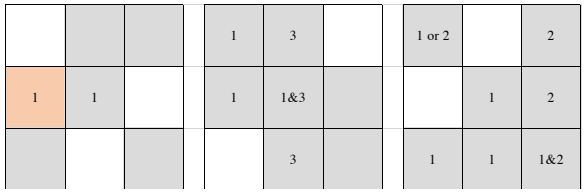
(18) The highlighted square can't have a 3 since it gives two different violations of backward consistency.



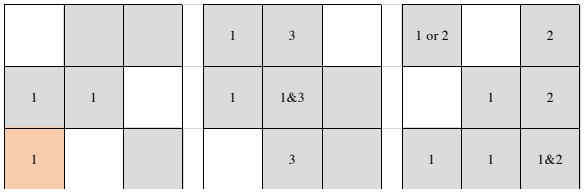
(19) The highlighted square can't have a 3 because either way we fill out (a,a,c) we get a violation of backward consistency. It can't have a 2 because of the immediate violation of backward consistency.



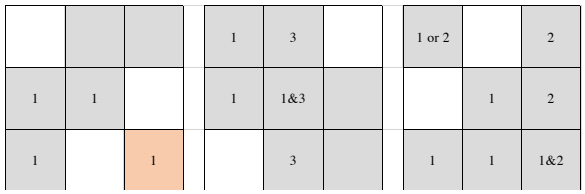
(20) 2 and 3 lead to immediate violations of backward consistency.



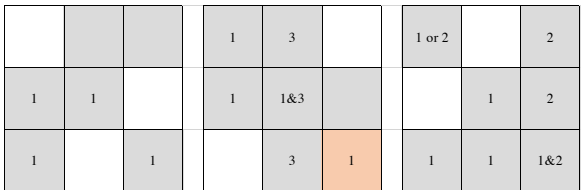
(21) 2 and 3 lead to immediate violations of backward consistency.



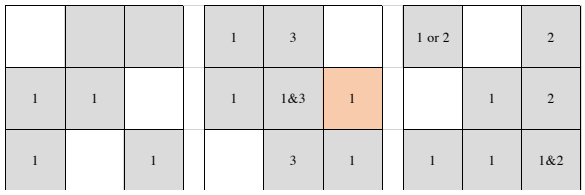
(22) 2 and 3 lead to immediate violations of backward consistency.



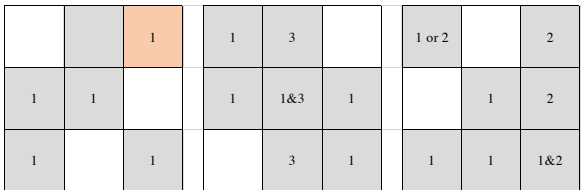
(23) 2 and 3 lead to immediate violations of backward consistency.



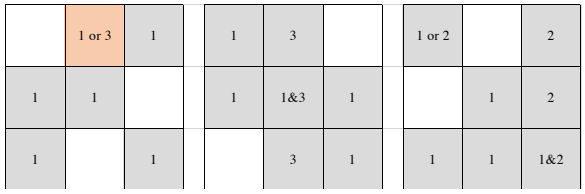
(24) 2 and 3 lead to immediate violations of backward consistency.



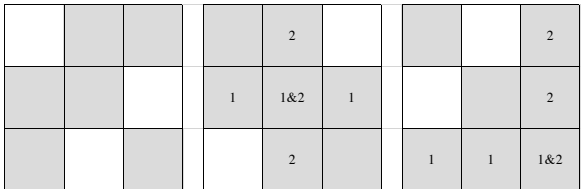
(25) 2 and 3 lead to immediate violations of backward consistency. The latter because we would need to label (b,a,b) 3 which already is labeled 1 and multiple labels cannot satisfy forward consistency.



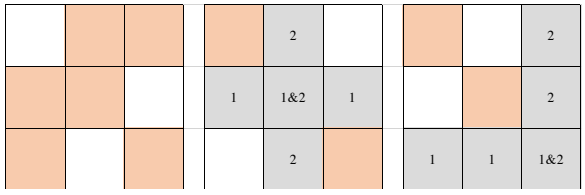
(26) 2 and 3 lead to immediate violations of backward consistency.



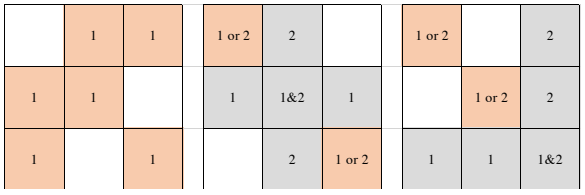
(27) Either way we fill out this last square does not lead to any violations.



(28) Finally we have the same pair compromise at both (b,b,b) and (c,c,c)



(29) None of the highlighted squares can be filled with a 3 since each would lead to a violation of backward consistency.



(30) The allocation in the second two panels can be independently assigned 1 or 2, while the first panel has a single equivalence class of T so all need to have the same label. By symmetry we simply choose 1

Chapter 3

Preface

Finally, I explore constrained allocation in the context of two-sided matching. Instead of pursuing mechanisms which are efficient and incentive compatible, I focus on mechanisms which produce a stable match. Again, the mechanism is limited to only producing outcomes which conform to an exogenous constraint. I introduce a mechanism, similar to deferred acceptance, except that agents choices are limited throughout the algorithm. While I focus on stability, I am still able to generate incentive compatibility with some additional restrictions.

Stable Matching Under General Constraints

Joseph Root[1]

Along with the increase in scope of applications for market design has come a need to develop greater context-specific flexibility. Whereas deferred acceptance, as introduced in Gale and Shapley (1962), is only concerned with the preferences of the agents to be matched, in practice, school districts, local governments, judges and others care about not only the outcome of any individual agent but about the structure of the outcome as a whole. As Roth (2002) points out, market designers don't have the luxury of ignoring these details. Consequently, market designers have introduced a number of diffuse solutions in the literature. This paper seeks to unify those solutions, both to provide an analytical framework for matching with general constraints, and to provide a simple model capable of providing flexibility to market designers facing real-world problems.

In school choice, diversity considerations are often mandated by law, but the precise requirement varies by city. In Boston, diversity requirements have been a part of the match process since 1974 when a judge ordered the city to achieve racial balance (Abdulkadiroglu 2013). In Jefferson County, the school district requires that schools maintain diversity by mandating schools admit students from low-income census tracts (Ehlers, Hafalir, Yenmez, and Yildirim 2013). New York requires schools to maintain balance of their student body by test score (Abdulkadiroglu 2013). However, the standard deferred acceptance algorithm from Gale and Shapley (1962) gives little guidance for how to implement DA in the presence of these constraints. Indeed, in the standard model, schools are assumed to have responsive [2] preferences which precludes non-trivial preferences over diversity. A series of solutions have emerged in the literature. Abdulkadiroglu and Sönmez (2003) introduced "controlled choice" as a means to achieve diversity. Controlled choice imposes upper bounds on the number of students at each school from a finite number of mutually exclusive groups (e.g. race, income quartiles, etc). For example, the school district might impose upper bounds on the number of students from a number of different neighborhoods. They showed that a

[1]Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: jroot@econ.berkeley.edu

[2]See Roth (1985) for a definition.

simple modification to deferred acceptance is compatible with these types of constraints (Abdulkadiroğlu 2005). More recently, Kojima (2012) showed that hard bounds can create inefficiency, even for students they intend to help. Ehlers, Hafalir, Yenmez, and Yildirim (2013) propose an alternative which treats bounds flexibly and they show the resulting matches Pareto dominate those from imposing hard bounds.

In residency matching, the geographic distribution of doctors has long been a central concern of governments and hospitals (Roth 1986). Residents compose a large proportion of many hospitals' labor force. Therefore, ensuring that a sufficient number of doctors are available to treat patients in each region is essential. However, the rural-hospitals theorem from Roth (1986) and Hatfield and Milgrom (2005) roughly states that the number of doctors matched with each hospital (and therefore each region) is the same for each stable match. This suggests a fundamental trade-off between stability and control over geographic diversity: if the matchmaker insists on stability of the match, they are stuck with the geographic distribution that results from DA, if instead they insist on control, there can be no guarantee of stability. Despite this apparent difficulty, the Japanese government decided to impose regional caps on the number of doctors assigned to any region. This gives rise to an immediate concern: how to distribute assign seats to hospitals within any one region. Kamada and Kojima (2015) propose a solution which they call flexible deferred acceptance. Specifically, they use an algorithm which imposes constraints for each hospital which depend on the number of applications each hospital has received. They then seek a limited form of stability in which the only blocking pairs are those which would lead to a violation of the regional capacity.[3] Kamada and Kojima (2019) provide a more general result which enables a rich hierarchy structure to the "regions" capable of modeling a wide class of constraints.

Motivated by the growing literature on application-specific constraints, this paper proposes a unified framework for studying matching problems with constraints. To do so, I introduce the notion of a constraint correspondence–a mapping which dynamically manages the choice of hospitals. For every subset of contracts, the correspondence offers each hospital a menu of choices, each of which is itself a subset of contracts satisfying the constraints. I then propose an algorithm called the constrained cumulative deferred acceptance algorithm which mimics traditional DA with the exception that hospital choices are constrained by the correspondence. The matching with contracts model of Hatfield and Milgrom (2005) corresponds to the special case in which the constraint correspondence is fully-flexible (in the sense that it imposes no restrictions). I show that so long as the constraint correspondence satisfies a condition I call "generalized substitutes," the constrained cumulative deferred acceptance algorithm will return a constrained stable outcome (a matching in which the only remaining blocking coalitions are those which would lead to a violation of the constraint). Generalized substitutes provides a set-valued generalization of the "substitutes condition" from Hatfield and Milgrom (2005).

[3]See Kamada and Kojima (2017b) for a discussion of the precise stability notion they use.

Strategy-proofness has proved to be an important desideratum in practical matching problems (Pathak and Sönmez 2008). In addition to the stability result, I provide two additional conditions on the constraint correspondence – aggregate monotonicity and constraint IIA – which ensure the existence of strategy-proof constrained stable mechanisms. Furthermore, I show that the constrained cumulative deferred acceptance algorithm is such a mechanism. Aggregate monotonicity is similar to the "law of aggregate demand" of Hatfield and Milgrom (2005) and ensures that as the set of available contracts expands, the constraints enable a weakly larger match. Constraint IIA places limits on the effect of an agent's reported preferences on the constraints of other agents. It ensures that no agent can costlessly announce a preference which does not affect her outcome directly, but that restricts the alternatives available to other agents.

Having developed this framework for matching with constraints, I then explore the relative efficiency of different constraint correspondences. Specifically, I provide a comparative statics result which enables comparison between mechanisms. Intuitively, I show that if one constraint correspondence is more permissive than another – in that hospitals choices from the first mechanism are always a subset of the choices from the second – then the latter makes all individuals on the *proposing* side better-off. I illustrate this result with an application to the flexible quotas of Ehlers, Hafalir, Yenmez, and Yildirim (2013) and show flexible quotas Pareto dominate inflexible quotas.[4]

The rest of the paper proceeds as follows: In the first section, I introduce notation and the basic definitions needed for the rest of the paper, develop the notion of constrained stability, and show that under the generalized substitutes condition a DA-type algorithm will yield a constrained stable outcome; in section 2, I prove a comparative statics result and demonstrate an application; in section 3, I provide conditions under which a constraint correspondence yields a strategy-proof mechanism. In the appendix I detail the connection between this model and the those of Kamada and Kojima (2019) and Ehlers, Hafalir, Yenmez, and Yildirim (2013).

3.1 Stability

Let D, H and X be disjoint finite sets which I will refer to as doctors, hospitals and contracts respectively. The use of the labels "doctors" and"hospitals" is simply convention. "Doctors" refers to the side of the market that will be matched with at most one contract, and "hospitals" to the side that may sign multiple contracts. In many settings it will be more appropriate to think of the two sides as "schools" and "students" or "firms" and "workers." The collection of contracts is equipped with functions $h : X \to H$ and $d : X \to D$ which specify that each contract $x \in X$ names one doctor $d(x)$ and one hospital $h(x)$. For example the set of contracts may be $D \times H$ so that each contract simply lists a doctor and hospital to be matched. In this case, $d(\cdot)$ and $h(\cdot)$ are the first and second coordinate maps, respectively. Alternatively, contracts might list the wage

[4]See theorem 8 in Ehlers, Hafalir, Yenmez, and Yildirim (2013) for their treatment of the Pareto comparison.

and other pertinent contractual details relevant to the participants. Given a subset of contracts $X' \subset X$ and z in D or H, write $X'_z = \{x \in X' : d(x) = z \text{ or } h(x) = z\}$ so that X'_z is the collection of contracts in X' which name z.

3.1.1 Preferences

Each doctor d is assumed to have strict preferences over contracts that name them, as well as an outside option which, by convention, I will denote $\emptyset$. That is, each d is associated with a linear order $\succ_d$ over $X_d \cup \{\emptyset\}$. If $\emptyset \succ_d x$ I will say that x is *unacceptable to d*. Hospital preferences are somewhat more complicated since they will be considering outcomes involving subsets of contracts. Each hospital h is assumed to have a physical capacity $q_h \geq 0$. This capacity is interpreted as the maximum number of contracts a hospital can sign [5]. For example, in the school choice setting, the capacity refers to the number of seats at each school. Not all subsets of contracts are possible for a hospital to sign. For example, no doctor may sign more than one contract. It will therefore be convenient to have an easy way to restrict attention to subsets of contracts to which hospitals could be matched. For each $h \in H$, and $X' \subset X$ let

$$\sigma_h(X') = \{Y \subset X'_h : |Y| \leq q_h \text{ and } x, y \in Y \text{ such that } x \neq y \implies d(x) \neq d(y)\}$$

Thus, fixing a hospital h, $\sigma_h(X')$ is simply the collection of subsets of contracts naming hospital h which do not exceed its capacity q_h and which do not include multiple contracts for the same doctor. Each hospital h is assumed to have a linear order on $\sigma_h(X)$ which I will denote $\succ_h$.

An **outcome** Y is a subset of contracts. Let $\mathcal{O} = 2^X$ denote the set of outcomes. Given hospitals and doctors preferences, it will sometimes be convenient to refer to the corresponding choice functions. For each $h \in H$, hospital h's choice function is the mapping $C_h : \mathcal{O} \to \mathcal{O}$ given by $C_h(X') = \max_{\succsim_h} \sigma_h(X')$. For each $d \in D$, doctor d's choice function is the mapping $C_d : \mathcal{O} \to X$ given by $C_d(X') = \max_{\succsim_d} [X'_d \cup \{\emptyset\}]$.[6] Equipped with these definitions, let $R_h(X') = X'_h - C_h(X')$ and $R_d(X') = X'_d - C_d(X')$. Note that R_h and R_d list only the doctors and hospitals, respectively which were rejected from the possible matches for each. That is, if X' contains a contract naming hospital $h' \neq h$, $R_h(X')$ does not include h', despite the fact that it was not chosen from X'.

Like the extant literature on two-sided matching, I will need to restrict hospital preferences to ensure limited complementarity between different contracts. In particular, I will assume the following condition on $\succ_h$ for each h.

Definition 7. The preference relation $\succ_h$ is **responsive**[7] on $\sigma_h(X)$ if

1. For all $X' \in \sigma_h(X)$ and $x, y \in X \setminus X'$ such that $X' \cup \{x\} \in \sigma_h(X)$ and $X' \cup \{y\} \in \sigma_h(X)$
$$X' \cup \{x\} \succ_h X' \cup \{y\} \iff \{x\} \succ_h \{y\}$$

[5] Since each doctor will sign at most one contract, this can also be interpreted as the upper limit on the number of doctors with whom a hospital can contract.

[6] Note that this definition implies that doctors choice functions have unit demand.

[7] This definition is motivated by the definition given in Roth (1985).

2. Whenever X', X'' are in $\sigma_h(X)$ and $X' \subset X''$ then $X' \prec_h X''$

The first condition requires that hospitals' preferences over singletons are context-free; if a hospital likes one contract more than another in isolation, it will still like the same contract when considering the same marginal tradeoff. The second condition says that all contracts are preferred to the outside option for hospitals: whenever a hospital can fill all its slots, it will do so. It is not difficult to weaken this condition by introducing an outside option for each hospital – much like I do for doctors – and assuming hospitals like sets of contracts only when all contracts are individually preferred to the outside option, but I choose not to do this to keep notation simple.

Readers familiar with the matching-with-contracts model of two-sided matching may wonder why I require that hospital preferences are responsive and have a capacity rather than the more familiar rejection monotonicity condition of Hatfield and Milgrom (2005):

$$X' \subset X'' \implies R_h(X') \subset R_h(X'') \text{ for all } h$$

The reason for the additional requirements is that rejection monotonicity is not enough to ensure substitutability over all choice problems a hospital might face. Rejection monotonicity only requires that the hospitals' choices does not exhibit complementarity over increasing sets. This is sufficient when the only choices hospitals need to make is over such sets (such as in the process of Deferred Acceptance). However, when constraining hospitals, one forces them to make choices over a larger collection of menus thereby increasing the space over which hospitals must have substitutable choices. The next example makes this point clear.

Example 1. Suppose that there are three contracts $\{x, y, z\}$ (all naming different doctors) and one hospital h with capacity 3. Let h's preferences be given by

$$\{x,y,z\} \succ_h \{x,z\} \succ_h \{x,y\} \succ_h \{y,z\} \succ_h \{x\} \succ_h \{y\} \succ_h \{z\} \succ_h \emptyset$$

Note then that

$$R_h(A) = \emptyset \text{ for all } A \in \mathcal{P}(X)$$

so $\succ_h$ satisfies the substitutes condition of Hatfield Milgrom (2005). These preferences (and their associated choice function), however, still exhibit a clear form of complementarity. Specifically

$$\{x,z\} \succ_h \{x,y\} \text{ and } \{y\} \succ_h \{z\}$$

so the presence of x flips the relative ranking of y and z. The intuition for why this will cause difficulties is the same as in Hatfield and Milgrom (2005) and will be discussed in more detail below.

3.1.2 The Constraints

Definition 8. Given a hospital $h \in H$, a **constraint correspondence for h** is a mapping $B_h : \mathcal{O} \to 2^{\mathcal{O}}$ such that $B_h(X') \subset \sigma_h(X')$ for all X'. A **constraint correspondence** is then a function $B = (B_h)_{h \in H} : \mathcal{O} \to \left(2^{\mathcal{O}}\right)^{|H|}$ such that each B_h is a constraint correspondence for h.

Intuitively, a constraint correspondence is a mapping that, for each subset of contracts, offers a menu of choices for each hospital. X' is informally thought of as the option set from which those choices may be drawn. The only restriction on B at this stage is that for each hospital and each option set, the menu consist of alternatives which satisfy the basic requirements that no hospital accept more than their quota and no hospital sign a single doctor more than once. The following examples are intended to illustrate this and later definitions. I will return to these examples throughout the paper and will discuss them more thoroughly in the applications section.

Running Examples.

- Consider B where $B_h(X') = \sigma_h(X')$ for all $h \in H$ and $X' \subset X$. I will call this the **fully-flexible** constraint correspondence. Notwithstanding the basic restrictions entailed by σ_h, it does not restrict hospital choices at all.

- On the other side of the spectrum, I can consider the class of constraint correspondences in which $B_h(X')$ is a singleton for each h and $X' \subset X$. That is, there is only one possible choice for each hospital for each collection of contracts. I will call this type of constraint correspondence **autocratic**.

- For a more substantive example, suppose that contracts are just doctor-hospital pairs so $X = D \times H$ and that the doctors can be partitioned into two types L and R. I think of L as"low-income" doctors and R as "rich" doctors. The government then wants to impose upper bounds on the number of rich doctors to whom a hospital can match. Formally, each h is given a quota $q_h^R \leq q_h$ and the constraint correspondence is given by $B_h(X') = \{Y \in \sigma_h(X') : |d(Y) \cap R| \leq q_h^R\}$.[8] Therefore at X' each hospital h is given its choice of any subset of contracts from $\sigma_h(X')$ which don't violate the cap.

- Now suppose that I are in the same situation as in the last example, but now the government lets hospitals exceed their quota of rich doctors only if there aren't enough low-income doctors in X' to fill the other seats. Formally, $B_h(X') = \{Y \in \sigma_h(X') : |Y \cap L| \geq \min\{q_h - q_h^R, |X'_h \cap L|\}\}$.

- Let $X = D \times H \times \{0,1\}$ where the 0-1 entry indicates whether the doctor gets a bonus. Assume that all doctors d like the bonus so that $(d,h,1) \succ_h (d,h,0)$ for all h. The central authority wants to use the bonus to incentivize doctors to match with under-demanded hospitals, so it only wants to offer the bonus if the hospital cannot fill its seats without it. To this end, let

$$B_h(X') = \{Y \in \sigma_h(X') : \text{if there is a } y \in Y \text{ s.t. } \tau_3(y) = 1, \tau_3(X'_h \setminus Y) = \{1\}\}$$

  where $\tau_3 : X \to \{0,1\}$ is the third coordinate projection (i.e. it returns 1 for a bonus contract and 0 otherwise). In words, $B_h(X')$ is all subsets in which h does not reject a bonus contract over a non-bonus contract.

[8] Notice that since $Y \in \sigma_h(X')$ each contract in Y names a distinct doctor, so $|d(Y)| = |Y|$.

Henceforth, I will fix a constraint correspondence B in order to avoid repeating that the definitions depend on the choice of B. I can define the hospitals' choices from B is a similar way to the definition of choice functions above.

Definition 9. For each $h \in H$, hospital h's **constrained choice function** is the mapping $\tilde{C}_h : \mathcal{O} \to \mathcal{O}$ given by $\tilde{C}_h(X') = \max_{\succsim_h} B_h(X')$. Similarly, for each $h \in H$, hospital h's **constrained rejection function** is the mapping $\tilde{R}_h : \mathcal{O} \to \mathcal{O}$ given by $\tilde{R}_h(X') = X'_h - \tilde{C}_h(X') = X'_h - \max_{\succsim_h} B_h(X')$.

It will also be convenient to be able to easily refer to "hospital-wide" and "doctor-wide" choices from a given subset. This motivates the following definition.

Definition 10. Define [9]

$$\tilde{C}_H(X') = \bigcup_{h \in H} \tilde{C}_h(X') \text{ and } C_D(X') = \bigcup_{d \in D} C_d(X')$$

and

$$\tilde{R}_H(X') = X' - \tilde{C}_H(X') \text{ and } R_D(X') = X' - C_D(X')$$

With the assumption that hospitals' preferences are responsive (as defined above), for any menu of choices, I can restrict attention to those that are not contained in another alternative choice.

Definition 11. A collection of contracts, $A \subset X$, is **maximal** at $B_h(X')$ if $A \in B_h(X')$ and there is no $A' \in B_h(X')$ such that A is a proper subset of A'. A is maximal at $B(X')$ if A is maximal at $B_h(X')$ for some h.

3.1.3 Constrained Stability

To motivate our notion of constrained stability, suppose that $X = D \times H$, there are four doctors and four hospitals with the following preferences:

d_1	d_2	d_3	d_4
h_1	h_3	h_1	h_1
h_2	h_4	$\vdots$	h_4
$\vdots$	$\vdots$		h_3
			$\vdots$
Low-income		rich	

h_1	h_2	h_3	h_4
$\{d_3, d_4\}$	$\{d_1\}$	$\{d_4\}$	$\{d_4\}$
$\{d_1, d_3\}$	$\vdots$	$\{d_2\}$	$\{d_2\}$
$\vdots$		$\vdots$	$\vdots$
Capacity 2	Capacity 1		

[9] We could equivalently define

$$\tilde{R}_H(X') = \bigcup_{h \in H} \tilde{R}_h(X') \text{ and } \tilde{R}_D(X') = \bigcup_{d \in D} R_d(X')$$

so that h_1 is d_1's favorite hospital, h_2 is her second favorite and so on. Doctors 1 and 2 are "low-income" and doctors 3 and 4 are "rich."

The principle desideratum of mechanisms in the two-sided matching literature is "stability" – the mechanism should find a match such that no coalition can leave the market to form a match amongst themselves which makes everyone involved better off. This goal is motivated by the empirical observation that unstable matches tend to perform badly in practice Roth and Sotomayor (1992). The two-sided matching literature since Gale and Shapley (1962) has furthermore shown that finding stable matches is always possible and there are efficient practical algorithms which can be used to find them.

In this particular setup, the matching

$$S_1 = \{(d_1, h_2), (d_2, h_3), (d_3, h_1), (d_4, h_1)\}$$

is stable – i.e. there are no blocking coalitions. In particular, everyone except d_1 is matched with their favorite partner(s). Note, however, that this match has h_1 accepting two rich doctors and rejecting a low-income applicant. In many settings, institutional and legal requirements enforce diversity constraints. For example, suppose I impose the requirement that h_1 can fill at most half of its seats with rich doctors. The matching

$$S_2 = \{(d_1, h_1), (d_2, h_3), (d_3, h_1), (d_4, h_4)\}$$

meets these requirements. Furthermore, in S_2 the low-income doctor d_1 is now matched with her top choice. However, in contrast to S_1, S_2 is not immune to coalitional deviations. In particular, d_3, d_4 and h_1 could deviate from S_2 and each be weakly better off – with h_1 and d_4 strictly better off. Note, however, that since d_3 and d_4 both prefer h_1 to all other hospitals and since h_1 similarly prefers $\{d_3, d_4\}$ to any other pair, unless they are matched together there will always be a coalitional deviation. Therefore, no match can simultaneously satisfy the legal requirements and be immune to deviation. It is easy to see, however, that S_2 does not have any additional coalitional deviations. Furthermore, the aforementioned deviation would lead to a violation of the constraints. This motivates a weaker notion of stability, related to the condition introduced in (Kamada and Kojima 2015): a match S' is **constrained stable** if it satisfies the constraints and if every blocking coalition would lead to a violation of the constraints. The idea is that a central authority – a school district or government, for example – has the ability to announce and reinforce restrictions. However, notwithstanding violations of their restrictions, the authority does not want to interfere. Therefore a constrained stable match is one in which for every possible coalitional deviation, the agents involved in the deviation know that they will be unable to maintain the block.

Unfortunately, policies aimed at helping low-income doctors can backfire. Suppose that instead of simply constraining hospital 1 to reserve seats for low-income doctors, I also constrain hospital 4 to not accept any rich doctors at all. Since S_2 has h_4 matched with d_4, it does not satisfy this requirement. However, the match

$$S_3 = \{(d_1, h_1), (d_2, h_4), (d_3, h_1), (d_4, h_3)\}$$

does. Like S_2, S_3 is subject to coalitional blocks. h_1, d_3 and d_4 could leave the match and make themselves better off. Furthermore, d_4 and h_4 could leave the match and make themselves better off. However, both would lead to violations of the diversity constraint so S_3 is constrained stable. One can quickly check that no other constrained stable match could make the doctors better off while respecting the constraints. Note that in S_3 both low-income doctors are weakly worse-off than in S_2, and d_2 is strictly worse-off. The attempt to strengthen the diversity constraints have backfired: both low-income doctors and rich doctors would have preferred a weaker policy.

There are ostensibly two culprits for this inefficiency:

1. The second set of constraints were poorly designed
2. Our notion of stability is too strict – I should only impose the bounds on rich students if loosening the bounds would help low-income students.

While the first source is indeed an issue, proper design would require substantial ex-ante knowledge about doctor preferences. An alternative profile of doctor preferences could have rendered the second set of constraints useful to the low-income doctors. For example, had it been that $h_4 \succ_{d_2} h_3$, the constraint would have ensured d_2 a spot at h_4 over d_4.

Instead, suppose that I alter our notion of constrained stability to require (1) either all hospitals respect their upper bound on rich doctors, or every hospital which violates its cap is not desired by any of the low-income doctors and (2) every blocking coalition would lead to a violation of the constraints. Using this definition, S_2, which was not constrained stable given the stronger constraints is now stable. To see why, notice that in S_2 both low-income doctors are matched with their favorite hospital. Hence, despite the fact that h_4 violates its cap of rich doctors, doing so does not preclude any low-income doctors from matching with h_4. Allowing the constraints to be flexible has led to an improvement for all doctors – rich and low-income.

Definition 12. Given $d \in D$ and $X' \subset X$, let $U_d(X')$ denote doctor d's upper contour set at X' [10] so that

$$U_d(X') = \begin{cases} \{x \in X : x \succsim_d y \text{ for some } y \in X'_d\} & \text{if } X'_d \neq \emptyset \\ X_d & \text{if } X'_d = \emptyset \end{cases}$$

Then define $U : \mathcal{O} \to \mathcal{O}$ by $U(X') = \bigcup_{d \in D} U_d(X')$.

If X' names each doctor at most once, then $U(X')$ is the set of contracts that doctors prefer to the outcome X'. Otherwise, it takes the upper contour set with respect to the worst contract in X' for each doctor. I may think of this as the "option set" available to hospitals when the proposed match is X'.

[10] That is, the set of contracts that d weakly prefers to her contract in X'. If X' contains, multiple contracts that name d, $U_d(X')$ will be the upper contour set for d with respect to d's least-favorite contract in X'.

Definition 13. An outcome X' is **feasible** if $X'_h \in B_h U(X')$ for all $h \in H$ and if X' names at most one contract for each doctor and if X_d is acceptable for each d. A collection of contracts X'' is said to **block** the outcome X' if $X'' \subset C_D(X' \cup X'')$ and for some h, $X'' \succ_h X'_h$ [11]. A collection of contracts X'' which blocks X' is said to **violate** B **at** X' if $X'' \notin B_h U(X')$.

A block is simply a hospital and collection of doctors which could profitably break away from the match and sign the blocking contracts to make everyone better off. The block violates the constraint if this set of contracts is not available to the associated hospital.

Definition 14. Given the constraint correspondence B, an outcome $X' \subset X$ is **constrained stable** if

(a) X' is feasible

(b) All blocking collections violate B at X'

Running Examples.

- In the case of the fully-flexible constraint correspondence, it is not difficult to verify that constrained stability reduces to the standard notion of stability from Hatfield and Milgrom (2005). Therefore for an outcome X' to be stable, there must be no blocking coalitions.

- Suppose B is autocratic and that X' is constrained stable. Then $X'_h = B_h U(X')$ for each h. Note that all blocking coalitions violate the constraint since there is only one choice for each hospital. Hence stability reduces to individual rationality.

- If B is given by $B_h(X') = \{Y \in \sigma_h(X') : |d(Y) \cap R| \leq q_h^R\}$ for all h, feasibility simply requires that X'_h does not exceed h's cap on rich doctors. A blocking coalition violates the constraint if it entails such a violation.

- If B is defined by $B_h(X') = \{Y \in \sigma_h(X') : |Y \cap L| \geq \min\{q_h - q_h^R, |X'_h \cap L|\}\}$, feasibility is more subtle. In particular, $X'_h \in B_h U(X')$ implies that X'_h exceeds h's capacity for rich doctors only if there are no low income doctors in $X'_h \backslash U(X')_h$. That is, if there are no low-income doctors who desire to matched with h at X' (i.e. they are either unmatched at X' or prefer h to their match in X'). A blocking coalition violates the constraint if it entails a violation of the cap.

- If $X = D \times H \times \{0, 1\}$ and

  $$B_h(X') = \{Y \in \sigma_h(X') : \text{if there is a } y \in Y \text{ s.t. } \tau_3(y) = 1, \tau_3(X'_h \setminus Y) = 1\}$$

  then feasibility implies that no hospital rejects a non-bonus contract in favor of a bonus contract. Then X' is constrained stable if the only possible deviations include subsets of contracts which entail a rejection of a non-bonus contract in favor of a bonus contract.

[11] Note that this implicitly requires that X'' only include contracts which name h.

The following lemma gives a simple alternative characterization of constrained stability. In particular, it says that an outcome X' is constrained stable if and only if X' is such that all hospitals would like to choose X'_h from their menu $B_hU(X')$, all contracts are acceptable to doctors, and at most one contract is named for each doctor.

Lemma 8. *The outcome $X' \subset X$ is constrained stable if and only if*

$$\tilde{C}_H U(X') = X' = C_D(X')$$

Equipped with this definition of stability, I might ask if one can guarantee the existence of a stable outcome for any constraint correspondence B. The following example gives a negative answer.

Example 2. Suppose that there are two doctors, d_1 and d_2, and two hospitals, h_1 and h_2, and that the set of contracts X is simply $D \times H$. Suppose doctors and hospitals have the following preferences:

$$\begin{gathered} h_1 \succ_{d_1} h_2 \\ h_2 \succ_{d_2} h_1 \\ \{d_1, d_2\} \succ_{h_1} \{d_1\} \succ_{h_1} \{d_2\} \succ_{h_1} \emptyset \\ \{d_1\} \succ_{h_2} \{d_2\} \succ_{h_2} \emptyset \end{gathered}$$

So h_1 and h_2 have physical capacities of 2 and 1, respectively. Suppose $B = (B_{h_1}, B_{h_2})$ where

$$\begin{aligned} B_{h_1}U(X') &= \{Y \subset \sigma_{h_1}U(X') : Y \neq \{d_1\}\} \\ B_{h_2}U(X') &= \{Y \subset \sigma_{h_2}U(X')\} \end{aligned}$$

In words, B constrains h_1 to not match with d_1 unless it can also match with d_2, in which case B requires that h_1 match with both. The following exhaustively demonstrates that no constrained stable matching exists.
The matches $\{(d_1, h_1)\}$, $\{(d_1, h_2), (d_2, h_2)\}$ and $\{(d_1, h_1), (d_2, h_2)\}$ are not feasible. The match $\{(d_2, h_1)\}$ is blocked by (d_1, h_2). $\{(d_1, h_2), (d_2, h_1)\}$ is blocked by (d_1, h_1). $\{(d_1, h_1), (d_2, h_1)\}$ is blocked by (d_2, h_2). $\{(d_2, h_2)\}$ is blocked by (d_1, h_2). $\{(d_1, h_2)\}$ is blocked by (d_2, h_1) (See figure 1).

This example works by imposing a complementarity in the constrained choices of hospital 1, despite no such complementarity in the underlying unconstrained choice function. The next two definitions will be helpful in guaranteeing that the constraint does not impose these types of complementarities.

Definition 15. A **monotone choice pair** for h is 4-tuple $(A', X', A'', X'') \in \mathcal{O}^4$ such that $X' \subset X''$ and A' is maximal at $B_h(X')$ and A'' is maximal at $B_h(X'')$.

Therefore a monotone choice pair is itself a pair of possible maximal choices from two subsets of contracts ordered by set inclusion. When $A \subset X$, $y \in A$ and $y' \in X \setminus A$ I will write $A_{y \to y'}$ as a shorthand for $[A \setminus \{y\}] \cup \{y'\}$.

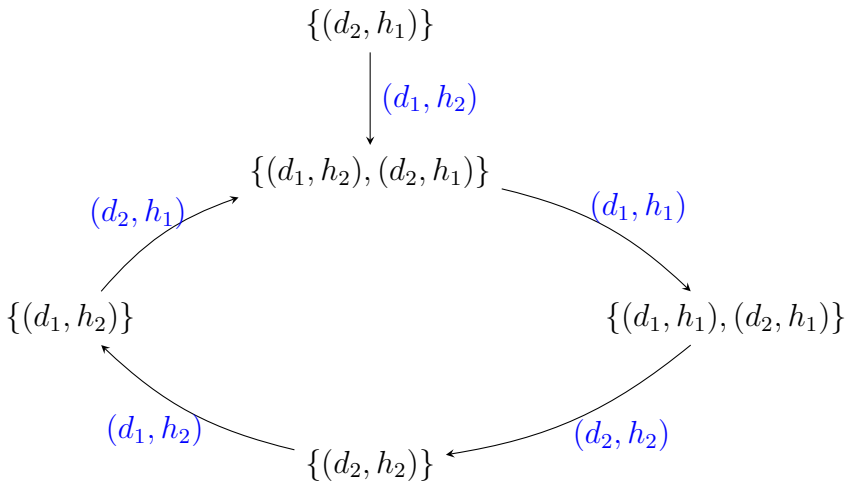


Figure 1. An Example with No Constrained Stable Outcome

Definition 16. The constraint correspondence B satisfies **generalized substitutes** [12] if for every h and every monotone choice pair for h, (A', X', A'', X'') such that there is a $y \in A'' \setminus A'$ with $y \in X'$, there is a $y' \in A' \setminus A''$ such that

$$A'_{y' \to y} \in B_h(X') \text{ and } A''_{y \to y'} \in B_h(X'')$$

Figure 2 demonstrates this property schematically. X' and X'' are in black, A' is in blue and A'' is in red. y is a contract available in X' which is not in A' (so if A' were chosen from $B_h(X')$, y would be rejected) but is in A''. The generalized substitutes condition then guarantees if such a point exists, there is another point y', also in X' which can be swapped for y' in both A' and A''.

Running Examples.

- It is not difficult to see that the fully-flexible constraint correspondence satisfies generalized substitutes.[13]
- The autocratic constraint correspondence gives motivation to the name of this condition. If $B_h(X')$ is a singleton for each h and $X' \subset X$, B_h satisfies generalized substitutes if and only if whenever $X' \subset X''$, $X'_h - B_h(X') \subset X''_h - B_h(X'')$.[14] That is, if I think of $B_h(X')$ as a choice function, then it must satisfy the substitutes condition of Hatfield and Milgrom (2005).

[12] A weaker condition dubbed a "substitutability with ties" is introduced in Erdil and Kumano (2019), for matching with priority structures in school choice.

[13] Any time a contract $y \in X'$ is not in $A \in B_h(X')$ it is because $|A'| = q_h$ (since A' is maximal). If A'' does contain y, however, then there is a doctor with a contract in A' with no contract in A''. Letting y' be any contract naming that doctor such that $y' \in X'$ gives the desired result.

[14] Note that there is a slight abuse of notation here since $B_h(X')$ is an element of $\mathcal{PP}(X')$. I ignore this issue when there is little likelihood of confusion.

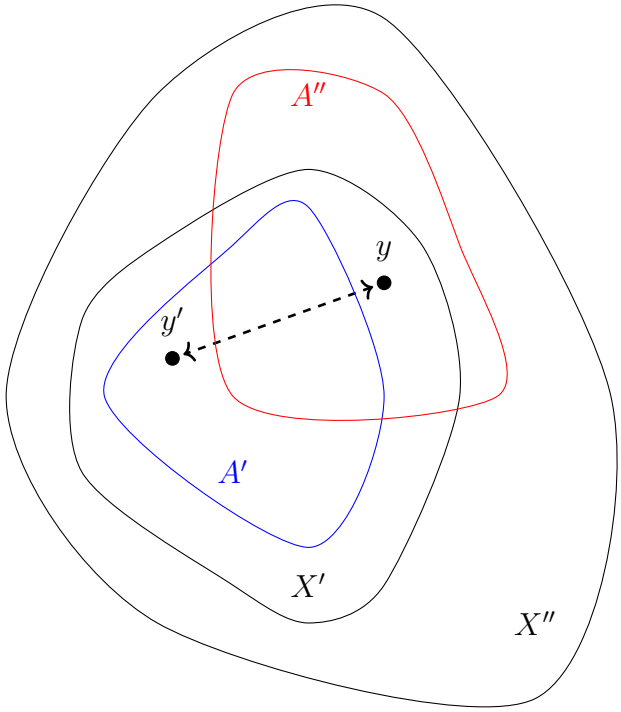


Figure 2. Generalized Substitutes

- The fact that the other two conditions satisfy the generalized substitutes condition is proved in the appendix.
- Let $X = D \times H \times \{0, 1\}$ and

$$B_h(X') = \{Y \in \sigma_h(X') : \text{if there is a } y \in Y \text{ s.t. } \tau_3(y) = 1, \tau_3(X'_h \setminus Y) = 1\}$$

and suppose (A', X', A'', X'') is a monotone choice pair. Suppose there is a $y \in X'$ with $y \in A'' \setminus A'$. If y is a bonus contract (so $\tau_3(y) = 1$), then $\tau_3(X'' \setminus A'')$. Hence A'' includes all non-bonus contracts in X'' so there are at most $q_h - 1$ such contracts in X''_h and therefore in X'_h. By responsiveness, and since A' is maximal and $y \notin A'$, $|A'| = q_h$. Hence there is at least one contract y' with $\tau_3(y') = 1$ in A' but not in A'' and the result follows. If instead $\tau_3(y) = 0$, A' and A'' both contain only non-bonus contracts and both have cardinality q_h. Since they aren't identical, the result is immediate.

The following lemma immediately demonstrates the benefit of this condition. It says that the constrained rejection function is monotone.

Lemma 9. *If the constraint correspondence B satisfies the generalized substitutes condition, whenever $X' \subset X'' \subset X$, $\tilde{R}_h(X') \subset \tilde{R}_h(X'')$ for all $h \in H$.* [15]

[15] This condition is referred to as the "substitutes" condition in Hatfield Milgrom (2005). It is also commonly referred to as "rejection monotonicity"

We are now ready to introduce a DA-type algorithm. This algorithm is intended to construct a constrained stable match.

Definition 17 (Constrained Cumulative Deferred Acceptance)**.**

Step 1 All doctors propose their favorite contract from X. Denote this set of contracts by X_1. Let each hospital h then choose their favorite subset from among $B_h(X_1)$, namely $\tilde{C}_h(X_1)$, and reject $\tilde{R}_h(X_1)$.

Step k Each doctor d proposes to her favorite contract not rejected in step $k-1$ as well as all contracts she weakly prefers to this contract. Denote this set of contracts by X_k. If $X_k = X_{k-1}$, terminate the algorithm and let $\tilde{C}_H(X_k)$ be the outcome. Otherwise, reject all contracts in $\tilde{R}_H(X_k) = \bigcup_{h \in H} \tilde{R}_h(X_k)$ and proceed to step $k+1$.

The following theorem establishes a positive result for any constraint correspondence satisfying the generalized substitutes condition. The proof closely follows Hatfield and Milgrom (2005).

Theorem 7. *If B satisfies generalized substitutability, the constrained cumulative deferred acceptance algorithm terminates in a finite number of steps and results in an outcome which is constrained stable. Furthermore, all doctors weakly prefer the outcome of the constrained cumulative differed acceptance algorithm to any other constrained stable outcome.*

3.2 Comparative Statics

Thus far I have introduced a family of constraint correspondences and have defined a DA-type algorithm for each. I have identified a subclass of constraint correspondences which can be guaranteed to yield a constrained stable outcome. However, the applied matchmaker might wonder how to find a constraint correspondence that is suitable to their problem. Furthermore, they may wonder if there is a way to compare two a priori acceptable constraint correspondences. This section provides an answer to this question. I develop a theorem which enables the matchmaker to compare constraint correspondences according to doctor preferences. In particular, I place a partial order $>$ on the class of constraint correspondences and show that if $B > B'$ then all doctors prefer the outcome of the constraint correspondence for B to that of B'. This is a partial order since not all constraint correspondences can be ordered in this way – some doctors may be made better off while others are made worse.

Theorem 8. *Suppose B and B' are constraint correspondences which satisfy the generalized substitutes condition. Suppose furthermore that for all $X' \subset X$ and $h \in H$*

whenever A and A' are maximal in $B_h(X')$ and $B'_h(X')$ respectively, if $y \in A \setminus A'$ then there is a $y' \in A' \setminus A$ such that

$$A'_{y' \to y} \in B'_h(X') \text{ and } A_{y \to y'} \in B_h(X')$$

then the outcome of constrained cumulative deferred acceptance for B' is weakly preferred by all doctors to the outcome of the constrained cumulative deferred acceptance algorithm for B.

To see a practical example of the usefulness of this result, consider the two constraint correspondences introduced in the running examples which correspond to limits on the number of "rich" doctors a hospital can hire. Recall they are defined by $B_h(X') = \{Y \in \sigma_h(X') : |Y \cap R| \leq q_h^R\}$ and $B'_h(X') = \{Y \in \sigma_h(X') : |Y \cap L| \geq \min\{q_h^R, |X'_h \cap L|\}\}$, respectively. The first imposes hard caps on the number of rich doctors a hospital can enroll, and the second imposes the same caps, but allows hospitals to violate the caps when they can't find enough low-income doctors to fill the remaining seats. The following corollary shows that the latter will not only make rich doctors better off (since they can now fill seats they were previously prohibited from filling), it will also make low-income doctors better off.

Corollary 2. *Suppose that for each $h \in H$, $0 \leq q_h^R \leq q_h$ and define B by $B_h(X') = \{Y \in \sigma_h(X') : |Y \cap R| \leq q_h^R\}$ and B' by $B'_h(X') = \{Y \in \sigma_h(X') : |Y \cap L| \geq \min\{q_h^R, |X'_h \cap L|\}\}$. Then, by theorem 2, the outcome of the constrained deferred acceptance algorithm for B' is weakly preferred by all doctors to the outcome of B.*

The following example furthermore shows that the soft constraints can, in fact, make the low-income doctors strictly better off.

Example 3. Suppose that there are two doctors and two hospitals, each with a capacity of one, with the following preferences:

d_1	d_2
h_1	h_2
h_2	h_1

h_1	h_2
$\{d_1\}$	$\{d_1\}$
$\{d_2\}$	$\{d_2\}$
$\emptyset$	$\emptyset$

and suppose that d_2 is low-income and d_1 is rich. Furthermore, suppose $q_{h_1}^R = 0$ and $q_{h_2}^R = 1$, so hospital 1 cannot enroll any rich doctors, whereas hospital 2 can admit either student. It's straightforward to see that the only constrained stable matching with respect to B is given by $S_1 = \{(d_1, h_2), (d_2, h_1)\}$. However, if instead I set $q_{h_2}^R = 0$ (or equivalently applied the fully-flexible constraint correspondence) the match $S^* = \{(d_1, h_1), (d_2, h_2)\}$ would be stable. This is clearly makes both doctors better off.

3.3 Incentives and Rural Hospitals

In this section I discuss two important results in the two-sided matching theory: the rural hospitals theorem and the strategy-proofness of deferred acceptance on the proposers' side. Both hold in our setting under suitable conditions.

Definition 18. We say that B satisfies **aggregate monotonicity** if whenever $X' \subset X'' \subset X$, for any maximal $(A_h)_{h\in H}$ in $B(X')$ and any maximal $(A'_h)_{h\in H}$ in $B(X'')$

$$\sum_{h\in H} |A_h| \leq \sum_{h\in H} |A'_h|$$

Theorem 9 (Rural Hospitals)**.** *If B satisfies aggregate monotonicity and generalized substitutability and if S and S' are two constrained stable matchings with respect to B then all doctors are either matched in both S and S' or are unmatched at both S and S'*

Thus to change the distribution of doctors, I will need to modify the constraint.

In practice, it is important that mechanisms don't incentivize agents to strategically misreport their preferences. The next condition is sufficient to guarantee that I can implement a constrained stable match as the outcome of a strategy-proof mechanism.

Definition 19. The constraint correspondence B satisfies **constraint IIA** if:
For any $X' \subset X$ and $(A_h)_{h\in H}$ such that for each $h \in H$, $A_h \in B_h(X')$ and A_h is maximal in $B_h(X')$, if $Y \subset X' - \bigcup_{h\in H} A_h$, then [16]

$$(A_h)_{h\in H} \in B(X' - Y) \subset B(X')$$

Theorem 10. *If B satisfies generalized substitutability, aggregate monotonicity and constraint IIA, the mechanism associated with the constrained cumulative deferred acceptance algorithm for B is strategyproof.*

3.4 Appendix

3.4.1 Proof of Lemma 8

Suppose that the outcome X' is constrained stable. By definition, $C_D(X') = X'$, so it sufficient to show that $X'_h = \tilde{C}_h U(X')$ for each h. For any $X'' \in B_h U(X')$, $X'' \subset C_D(X' \cup X'')$ since $X'' \subset U(X')$ and X'' names each doctor at most once. However, since X' is constrained stable, it must be that $X'_h \succsim_h X''$, since otherwise X'' would block X' and would not violate the constraint. This shows that $X'_h = \max_h B_h U(X') = \tilde{C}_h U(X')$ which was the desired result. Conversely, suppose that $\tilde{C}_H U(X') = X' = C_D(X')$. Clearly, X' is feasible. Now suppose that X'' blocks X' so that $X'' \succ X'_h$ for some h and $X'' \subset C_D(X' \cup X'')$. Then since $X'_h = \max_h B_h U(X')$ it cannot be that $X'' \in B_h U(X')$. Hence X'' violates B at X'. Since X'' was chosen arbitrarily, all blocking collections violate B at X'. □

[16] Set inclusion refers to set inclusion of each element of the product.

3.4.2 Proof of Lemma 9

Suppose $y \in \tilde{R}_h(X')$ and $y \in A'' \in B_h(X'')$. I will show that $A'' \neq \tilde{C}_h(X'')$. First, if A'' is not maximal, I are done. If instead, A'' is maximal, the 4-tuple $(\tilde{C}_h(X'), X', A'', X'')$ is a monotone choice pair. If B satisfies generalized substitutes, there is a $y' \in \tilde{C}_h(X') \setminus A''$ such that

$$\left[\tilde{C}_h(X') \setminus \{y'\}\right] \cup \{y\} \in B_h(X') \text{ and } [A'' \setminus \{y\}] \cup \{y'\} \in B_h(X'')$$

Then by definition, $\tilde{C}_h(X') \succ_h \left[\tilde{C}_h(X') \setminus \{y'\}\right] \cup \{y\}$. However, since hospital preferences are responsive,

$$\tilde{C}_h(X') \succ_h \left[\tilde{C}_h(X') \setminus \{y'\}\right] \cup \{y\} \iff \{y'\} \succ_h \{y\} \iff [A'' \setminus \{y\}] \cup \{y'\} \succ_h A''$$

Hence $A'' \neq \tilde{C}_h(X'')$. Therefore, $\tilde{C}_h(X'')$ and $\tilde{R}_h(X')$ are disjoint and $\tilde{R}_h(X'') \supset \tilde{R}_h(X')$ as desired. □

3.4.3 Proof of Theorem 7

Let $F_1 : \mathcal{O} \to \mathcal{O}$ be defined by $F_1(X') = X - \tilde{R}_H(X')$ and $F_2 : \mathcal{O} \to \mathcal{O}$ be defined by $F_2(X') = X - R_D(X')$ and let $F : \mathcal{O} \times \mathcal{O} \to \mathcal{O} \times \mathcal{O}$ by

$$F(X_1, X_2) = [F_1(X_2), F_2F_1(X_2)]$$

Define the partial order $\geq$ on $\mathcal{O} \times \mathcal{O}$ by

$$(X_1, X_2) \geq (X_1', X_2') \iff X_1 \supset X_1' \text{ and } X_2 \subset X_2'$$

Now, if $X_2 \subset X_2'$, by lemma 9, $\tilde{R}_h(X_2) \subset \tilde{R}_h(X_2')$ hence $\tilde{R}_H(X_2) \subset \tilde{R}_H(X_2')$ and

$$F_1(X_2) = X - \tilde{R}_H(X_2) \supset X - \tilde{R}_H(X_2') = F_1(X_2')$$

Since doctors have unit-demand, if $X' \subset X'' \subset X$, $R_D(X') \subset R_D(X'')$. Hence $F_1(X_2) \supset F_1(X_2')$ implies that $R_D[F_1(X_2)] \supset R_D[F_1(X_2')]$ Thus

$$F_2F_1(X_2) = X - R_D[F_1(X_2)] \subset X - R_D[F_1(X_2')] = F_2F_1(X_2')$$

Together, this shows that $(X_1, X_2) \geq (X_1', X_2')$ implies $F(X_1, X_2) \geq F(X_1', X_2')$. Therefore, F is order-preserving on the complete lattice $(\mathcal{O} \times \mathcal{O}, \geq)$. By the Knaster-Tarski fixed-point theorem, the set of fixed points of F is a complete nonempty lattice. Furthermore, iteratively applying F to $(X, \emptyset)$ gives the largest fixed point.

Now suppose that S is a constrained stable outcome and consider

$$(S \cup [X \setminus U(S)], U(S))$$

By lemma 8, $X - \tilde{R}_H(U(S)) = X - (U(S) - S) = S \cup [X \setminus U(S)]$ Next, observe that $S \cup [X \setminus U(S)]$ is the union of doctors weak lower contour set at S. Hence $R_D(S \cup [X \setminus U(S)]) = X \setminus U(S)$ and $X - R_D(S \cup [X \setminus U(S)]) = U(S)$. This shows that $(S \cup [X \setminus U(S)], U(S))$ is a fixed point of F.

Finally, it remains to show that the largest fixed point of F corresponds to the outcome of the constrained cumulative deferred acceptance algorithm and that it is constrained stable. Starting with $(X, \emptyset)$ and iteratively applying F corresponds exactly to the steps outlined in words in the constrained cumulative deferred acceptance algorithm. At each step, doctors propose all previously proposed contracts and potentially a new contract [17] Hospitals, consider this collection and choose from their constrained choice set, rejecting all other contracts proposed to them. Let $\left(\bar{X}_1, \bar{X}_2\right)$ be the largest fixed point of F. Then the constrained cumulative deferred acceptance algorithm produces $\tilde{C}_H(\bar{X}_2)$ as the proposed outcome. For notational convenience, let $\bar{S} = \tilde{C}_H(\bar{X}_2)$. Then by the above argument, $\bar{X}_2 = U(\bar{S})$. Furthermore, since $\left[\bar{X}_1, \bar{X}_2\right]$ is a fixed point of F, $\bar{X}_1 = X - (U(\bar{S}) - \bar{S})$ and $U(\bar{S}) = X - \tilde{R}_H(\bar{X}_1)$, so by lemma 8, $\bar{S}$ is constrained stable.

If the outcome S is constrained stable,

$$(S \cup [X \setminus U(S)], U(S)) \leq \left(\bar{S} \cup \left[X \setminus U(\bar{S})\right], U(\bar{S})\right)$$

In particular, $U(S) \supset U(\bar{S})$, so all doctors are weakly better of at $\bar{S}$.

□

3.4.4 Proof of Theorem 8

Part 1: $\max_{\succsim_h} B_h(X') \subset \max_{\succsim_h} B'_h(X')$ for all $h \in H$ and $X' \subset X$ [18]

Fix $X' \subset X$ and $h \in H$. Suppose that $y \in X' \setminus \max_{\succsim_h} B'_h(X')$. Then for any maximal A in $B_h(X')$, if $y \in A$, by the hypothesis, there is a y' such that $y' \in \max_{\succsim_h} B'_h(X')$ and such that

$$\left[\max_{\succsim_h} B'_h(X') \setminus \{y'\}\right] \cup \{y\} \in B'_h(X')$$

However, by definition $\max_{\succsim_h} B'_h(X') \succ_h [\max_{\succsim_h} B'_h(X') \setminus \{y'\}] \cup \{y\}$. Since hospital preferences are responsive, this implies that $\{y'\} \succ_h \{y\}$ and $[A \setminus \{y\}] \cup \{y'\} \succ_h A$. Furthermore, $[A \setminus \{y\}] \cup \{y'\}$ is in $B_h(X')$, so $A \neq \max_{\succsim_h} B_h(X')$. This shows, more generally, that if $A \in B_h(X')$ and if $A \setminus \max_{\succsim_h} B'_h(X') \neq \emptyset$ then $A \neq \max_{\succsim_h} B_h(X')$. Hence $\max_{\succsim_h} B_h(X') \subset \max_{\succsim_h} B'_h(X')$.

Part 2: All doctors weakly prefer the outcome of the constrained cumulative deferred acceptance algorithm for B' to the outcome of the constrained cumulative deferred acceptance algorithm for B.

[17] If they were rejected in the previous step and still have contracts they have not yet proposed.

[18] We are using the cumbersome notation $\max_{\succsim_h} B_h(X')$ instead of $\tilde{C}_h(X')$ since the latter is defined with respect to a fixed constraint correspondence. In this theorem, I are explicitly considering two different constraint correspondences, so I use the more difficult notation to maintain clarity.

As in the proof of theorem 7,

$$F_1(X') = X - \left[X' - \bigcup_{h\in H} \max_{\succsim_h} B_h(X')\right] \text{ and } F_2(X') = X - [X' - C_D(X')]$$

and

$$F_1'(X') = X - \left[X' - \bigcup_{h\in H} \max_{\succsim_h} B_h'(X')\right] \text{ and } F_2'(X') = F_2(X')$$

Then let $F(X_D, X_H) = [F_1(X_H), F_2F_1(X_H)]$ and $F'(X_D, X_H) = [F_1'(X_H), F_2'F_1'(X_H)]$. By the proof of the main theorem, if S is the outcome of constrained cumulative deferred acceptance for B, then $[S \cup [X - U(S)], U(S)]$ is a fixed point for F. By part 1, $\max_{\succsim_h} B_h(U(S)) \subset \max_{\succsim_h} B_h'(U(S))$ for all h so

$$F_1'(U(S)) = X - \left[U(S) - \bigcup_{h\in H} \max_{\succsim_h} B_h'(U(S))\right] \supset X - \left[U(S) - \bigcup_{h\in H} \max_{\succsim_h} B_h(U(S))\right]$$

and

$$X - \left[U(S) - \bigcup_{h\in H} \max_{\succsim_h} B_h(U(S))\right] = F_1(U(S)) = S \cup [X - U(S)]$$

together this gives that $F_1'(U(S)) \supset F_1(U(S)) = S \cup [X - U(S)]$ but then

$$F_2'F_1'(U(S)) \subset F_2'F_1(U(S)) = F_2F_1(U(S)) = U(S)$$

Finally, this shows that $F'[S \cup [X - U(S)], U(S)] \geq [S \cup [X - U(S)], U(S)]$ where " $\geq$ " is the partial order on $\mathcal{P}(X)^2$ introduced in the proof of theorem 7. Furthermore, as established in the proof of theorem 7, F' is order-preserving, so F' converges to a fixed point weakly larger than $[S \cup [X - U(S)], U(S)]$. Since the constrained cumulative deferred acceptance algorithm for B' produces the largest fixed point, I get the desired result. □

3.4.5 Proof of Theorem 9

Suppose S^* is the constrained stable outcome of the constrained cumulative deferred acceptance algorithm for B. Then all doctors are weakly worse off at S and S' by the proof of theorem 7 above. Thus it must be that $|S_d| \leq |S_d^*|$ and that $|S_d'| \leq |S_d^*|$ for all d. Summing these up, I get

$$|S| = |\cup_{d\in D} S_d| = \sum_{d\in D} |S_d| \leq \sum_{d\in D} |S_d^*| = |S^*|$$

and

$$|S'| = |\cup_{d\in D} S_d'| = \sum_{d\in D} |S_d'| \leq \sum_{d\in D} |S_d^*| = |S^*|$$

Furthermore, $U(S) \supset U(S^*)$ and $U(S') \supset U(S^*)$. Hence by aggregate monotonicity $|S| \geq |S^*|$ and $|S'| \geq |S^*|$. Together with the above inequalities, I then have that $|S| = |S^*| = |S'|$. Therefore, $|S_d| = |S_d^*|$ and $|S_d'| = |S_d^*|$, since if any of these inequalities were strict, I would violate the above. □

3.4.6 Proof of Theorem 10

Fix B and the preferences of hospitals and consider some d and preference profile $\succsim = (\succsim_d)_{d \in D}$ where

$$x_1 \succ_d \cdots \succ_d x_m$$

Let S be the outcome of the constrained cumulative deferred acceptance algorithm for $\succsim$ and let $S_d = x_j$. Consider $\succsim_d'$ given by

$$x_j \succ_d' x_1 \succ_d' \cdots \succ_d' x_{j-1} \succ_d' x_{j+1} \succ_d' \cdots \succ_d' x_m$$

and let $U(S)$ be defined with respect to $\succsim$ and $\hat{U}(S)$ be defined with respect to $(\succsim_d', \succsim_{-d})$ so that $\hat{U}(S) = U(S) \setminus \{x_1, x_2, \dots x_{j-1}\}$. By lemma 8, S is constrained stable under $(\succsim_d', \succsim_{-d})$. Hence d is matched with x_j at the constrained cumulative deferred acceptance outcome for $(\succsim_d', \succsim_{-d})$.

Now consider $\succsim_d^*$ given by

$$y_1 \succ_d^* y_2 \succ_d^* \cdots \succ_d^* x_j \succ_d^* \dots y_{m-1}$$

and let S^* be the outcome of the constrained cumulative deferred acceptance algorithm given the preference profile $(\succsim_d^*, \succsim_{-d})$. Let $U(S^*)$ be defined with respect to $(\succsim_d', \succsim_{-d})$ and $\hat{U}(S^*)$ be defined with respect to $(\succsim_d^*, \succsim_{-d})$. Suppose that $S_d^* = \emptyset$. Then $U(S^*) = \hat{U}(S^*)$, so S^* is constrained stable under $(\succsim_d', \succsim_{-d})$ by lemma 8. However this contradicts theorem 9, so it must be that $S_d^* \neq \emptyset$. Suppose that $x_j \succ_d^* S_d^*$ then by constraint IIA, if d were to submit

$$x_j \succ_d^* S_d^* \succ_d \dots$$

she would still be assigned S_d^*. However, this contradicts the conclusion above that if x_j were at the top of d's list, she would achieve it. Therefore, $S_d^* \succ_d^* x_j$ which gives the result. □

3.5 School Choice Constraints With Multiple Types

3.5.1 Controlled Choice with Hard Upper Bounds

For this example, let $X = D \times H$, so each contract simply specifies a single doctor to be matched with a single hospital. Suppose that doctors are partitioned into a finite number of types $T_1, T_2, \dots, T_n$ which specify characteristics over which the matchmaker

(e.g. the school district or local government) would like to achieve diversity. For example, in the school choice setting, types might be socioeconomic indicators or racial categories. In order to achieve diversity, the matchmaker imposes hard upper bounds[19] on the number of doctors from each type a hospital can enroll. Formally, for each $h \in H$ there is a vector $\bar{q}_h = (q_h^i)_{i=1}^n \geq 0$. Each hospital can admit at most q_h^i doctors from type T_i. Consider the constraint correspondence for h

$$B_h(X') = \{Y \in \sigma_h(X') : |d(Y) \cap T_i| \leq q_h^i \text{ for } i = 1, \ldots, n\}$$

and the associated constraint correspondence $B = (B_h)_{h \in H}$. This constraint attempts satisfy the bounds in a straightforward way: hospitals are never allowed to exceed their upper bound from each type. Notwithstanding this limitation, B enables h to choose freely from among the all subsets of doctors which satisfy the basic requirements of σ.

Proposition 11. *The constraint correspondence*

$$B_h(X') = \{Y \in \sigma_h(X') : |d(Y) \cap T_i| \leq q_h^i \textit{ for } i = 1, \ldots, n\}$$

satisfies generalized substitutes, aggregate monotonicity and constraint IIA.

Proof.

1. Generalized Substitutes

   Suppose that (A', X', A'', X'') is a monotone choice pair and that $y \in A'' \setminus A'$ and $y \in X'$. Let i be the index such that $d(y) \in T_i$. Since $y \notin A'$, and A' is maximal, either $|A'| = q_h$ or $|A' \cap T_i| = q_h^i$ (otherwise, I would be able to add y to A without violating the upper bounds). I will first consider case 1. If $|A'| = q_h$, $|A''|$ must also be q_h since $|A''| < |A'|$ implies that for some j, $|A'' \cap T_j| < |A' \cap T_j|$. In which case, for any $z \in [A' \setminus A''] \cap T_j$ the collection $A'' \cup \{z\} \in B_h(X'')$ so A'' is not maximal. Hence $|A''| = q_h$. It will be useful to break case 1 into two additional cases. First, if $[A' \setminus A''] \cap T_i \neq \emptyset$ then for any $y' \in [A' \setminus A''] \cap T_i$ I the desired statements

   $$[A' \setminus \{y'\}] \cup \{y\} \in B_h(X') \text{ and } [A'' \setminus \{y\}] \cup \{y'\} \in B_h(X'')$$

   Second, if $[A' \setminus A''] \cap T_i = \emptyset$ then there is a $j \neq i$ such that $|A' \cap T_j| > |A'' \cap T_j|$ and for any $y' \in [A' \setminus A''] \cap T_j$ I get the same statements. Next, consider case 2 in which $|A' \cap T_i| = q_h^i$. Then $[A' \setminus A''] \cap T_i$ is nonempty and for any $y \in [A' \setminus A''] \cap T_i$ I get the desired result.

2. Aggregate Monotonicity

   Suppose that $X' \subset X''$ and fix $h \in H$. Let A' be a maximal element of $B_h(X')$ and let $A'' \in B_h(X'')$. If $|A''| < |A'|$ then for some i, $|A'' \cap T_i| < |A' \cap T_i| \leq q_h^i$ in which case there is a $y \in [A' \setminus A''] \cap T_i$. Hence $A'' \cup \{y\} \in B_h(X'')$ so A'' is not maximal. This establishes that if A'' is maximal in $B_h(X'')$ then $|A'| \leq |A''|$. Summing over hospitals, I get the desired result.

[19] See Abdulkadiroğlu (2005) for a more thorough discussion on this topic.

3. Constraint IIA

   Suppose that $X' \subset X$ and $(A_h)_{h\in H}$ is a vector of maximal choices from $B_h(X')$ Suppose that $Y \subset X' - \bigcup_{h\in H} A_h$ and consider $B(X' - Y)$. First, the upper bounds haven't changed so $A_h \in B_h(X' - Y)$ for each h. Second, any element of the $B_h(X')$ still respects the upper bounds when Y is removed from X'.

□

Corollary 3. *The constrained deferred acceptance algorithm for B yields a constrained stable outcome which is weakly preferred by all doctors to any other constrained stable outcome. Furthermore, the mechanism associated to B is strategy-proof for doctors.*

Proof. This is an immediate consequence of proposition 11 and theorems 7 and 10. □

3.5.2 Controlled Choice with Soft Bounds

As discussed in Ehlers, Hafalir, Yenmez, and Yildirim (2013) and Kojima (2012), the imposition of hard bounds can come with a cost. This inefficiency can be remedied by treating bounds as *soft*. That is, the bounds only bite when a hospital has reached their capacity – at which point the bounds dictate which tradeoffs hospitals can make by type. Following Ehlers, Hafalir, Yenmez, and Yildirim (2013), I can impose both upper and lower bounds. As detailed in the last section, upper bounds place a maximum on the number of doctors from each type hospitals can admit. By contrast, lower bounds place a limitation on the minimum number of doctors that each hospital can admit from a given type. In order to accomodate this type of constraint, B will ensure a dynamic priority of doctors. First, when possible, B will ensure that the lower bounds will be met. When not possible say for type T_i, B will ensure that hospitals admit all doctors from type T_i who desire to match with it. Second, hospitals will be required to admit all doctors from types which have met their floors but not yet exceeded their ceilings. Only after satisfying these two requirements can hospitals exceed their upper bound for a given type.

Let $X = D \times H$. Again suppose that doctors can be partitioned into types $T_1, \ldots, T_n$. I will abuse notation slightly and write $y \in T_i$ for $y \in X$ to mean that $d(y) \in T_i$. The matchmaker imposes upper and lower bounds on the number of doctors from each type for each hospital. That is, there are vectors $\underline{q}_h = (\underline{q}_h^i)_{i=1}^n \geq 0$ and $\bar{q}_h = (\bar{q}_h^i)_{i=1}^n \geq 0$ for each h such that $\underline{q}_h \leq \bar{q}_h$. I will assume that $\sum_{i=1}^n \underline{q}_h^i \leq q_h$ for each h.

In order to find the right statement of the constraint correspondence, it will be useful to first establish a bit of notation. For $h \in H$ and $X' \subset X$, let

$$P_h^1(X') = \{Y \subset \sigma_h(X') : |Y \cap T_i| \geq \min\{\underline{q}_h^i, |X'_h \cap T_i|\} \text{ for } i = 1, 2, \ldots, n\}$$

therefore, $P_h^1(X')$ is the collection of subsets of contracts which either meet each of the lower bounds for each type or, if not possible, then include all doctors. Note that this

is only part of the way to the constraints that I would like to impose: conditional on meeting the lower bounds, this allows hospitals to exceed the upper bounds, even when they could allocate more seats to types that have not yet met the upper bound.

Now, let

$$P_h^2(X') = \{Y \subset \sigma_h(X') : \exists i, j \text{ s.t. } |Y \cap T_i| > \bar{q}_h^i \text{ and } |Y \cap T_j| < \min\{|X'_h \cap T_i|, \bar{q}_h^i\}\}$$

these are the types of subsets of doctors which entail h admitting more than its upper cap of some type i, while admitting fewer than its upper cap of another type j, despite the fact that X'_h contains doctors of type j sufficient to reach j's cap. These are the types of subsets I would like B to rule out. Our goal is to first prioritize all doctors up to the lower bound, then to prioritize those who are above the lower bound, but below the upper bound, and only last to prioritize doctors who exceed the upper bound. collections of doctors in $P_h^2(X')$ violate this hierarchy.

Having done this notational work, I can now easily define B as follows:

$$B_h(X') = P_h^1(X') - P_h^2(X')$$

Let us now examine constrained stability given B. Recall that the three requirements for an outcome X' to be stable are (1) X' is feasible (2) $C_D(X') = X'$ and (3) all blocking coalitions violate the constraint. I will examine the implications of each. First, if $X'_h \in B_h U(X')$ then $|X'_h \cap T_i| \geq \min\{\underline{q}_h^i, |U(X')_h \cap T_i|\}$ for each i. Hence, for i such that $|U(X')_h \cap T_i| \geq \underline{q}_h^i$, X'_h entails enrollment of at least $\underline{q}_h^i$ doctors from type i. For i such that $|U(X')_h \cap T_i| < \underline{q}_h^i$, X'_h includes all doctors from type i in the upper contour set at X'. Furthermore, $X'_h \in B_h U(X')$ implies that if there is any i such that $|X'_h \cap T_i| > \bar{q}_h^i$, it must be that there is no j such that $|X'_h \cap T_j| < \min\{|U(X')_h \cap T_i|, \bar{q}_h^i\}$. In words, this states that B only allows j to exceed its upper bound for any type if not doing so for all types would lead to vacant seats. Next, the condition that $C_D(X') = X'$ is simply a regularity condition enforcing that X' does not specify that a single doctor be matched to multiple hospitals. Finally, suppose that the first two conditions for constrained stability are satisfied by X'. Now, suppose that X'' is a blocking coalition (so that $X'' \subset C_D(X' \cup X'')$ and there is a h such that $X'' \succ_h X'_h$. This blocking coalition violates the constraint if $X'' \notin B_h U(X')$. Again, note that $X'' in U(X')$ by definition. Therefore $X'' \notin B_h U(X')$ implies that either X'' is not in $P_h^1 U(X')$ or X'' is in $P_h^2 U(X')$. The former implies that X'' would entail that h enroll fewer than the lower bound for some type i, despite $|U(X') \cap T_i| > \underline{q}_h^i$. The latter implies that h admitting X'' would exceed its upper bound for some type of doctor while there are still doctors who would like to enroll with h from other types that have not yet reached their bound.

The following proposition establishes that our theory is compatible with soft bounds.

Proposition 12. *The constraint correspondence $B_h(X') = P_h^1(X') - P_h^2(X')$ satisfies generalized substitutes.*

Proof. Suppose that (A', X', A'', X'') is a monotone choice pair and that $y \in A'' \setminus A'$ and $y \in X'$. Let i be the index such that $d(y) \in T_i$. First, I will show that $|A'| = q_h$. Since A' is maximal, $A' \cup \{y\} \notin B_h(X')$. Suppose that $|A' \cup \{y\}| \leq q_h$ so $A' \cup \{y\} \in P_h^1(X')$ since A' is in $P_h^1(X')$. Hence it must be that $A' \cup \{y\} \in P_h^2(X')$, despite the fact that $A' \in P_h^2(X')$. Clearly then there is a j such that $|A' \cup \{y\} \cap T_j| < \min\{|X'_h \cap T_i|, \bar{q}_h^i\}$ however since $|A' \cup \{y\}| \leq q_h$ by assumption, $|A'| < q_h$. Thus for any $z \in [X'_h \setminus A'] \cap T_i$, $A' \cup \{z\}$ is in $B_h(X')$. This contradicts the fact that A' is maximal. hence it must be that $|A' \cup \{y\}| > q_h$, so $|A'| \geq q_h$. Together with the fact that $A' \in \sigma_h(X')$ I get $|A'| = q_h$. Next, I will establish that $|A''| = q_h$. However, since $|A''| \leq q_h$, $A' \setminus A''$ is nonempty. Following the same argument above with A'' and any element of $A' \setminus A''$ will establish the desired result. Hence $|A'| = q_h = |A''|$.

Having established the number of contracts in A' and A'', it will be useful to break the problem down into three cases. First, suppose that $|A' \cap T_i| = |A'' \cap T_i|$. Then clearly there is a $y' \in A' \cap T_i$ such that $y' \notin A''$. However swapping out two contracts from the same type will not change the constraints from B. Therefore,

$$[A' \setminus \{y'\}] \cup \{y\} \in B_h(X') \text{ and } [A'' \setminus \{y\}] \cup \{y'\} \in B_h(X'')$$

Next, suppose that $|A' \cap T_i| < |A'' \cap T_i|$ then there is a $j \neq i$ such that $|A' \cap T_j| > |A'' \cap T_j|$. Let y' be a contract in $[A' \setminus A''] \cap T_j$. Consider swapping y' for y in A'. Since $|A' \cap T_j| > |A'' \cap T_j|$ and $y' \in X''$, $|A' \cap T_j| > \underline{q}_h^j$, so the swap will not cause issues with the lower bound, so $[A' \setminus \{y'\}] \cup \{y\} \in P_h^1(X')$. It remains to show that $[A' \setminus \{y'\}] \cup \{y\} \notin P_h^2(X')$. Since $A' \in P_h^2(X')$ either there is a k such that $|A' \cap T_k| > \bar{q}_h^k|$ in which case for all $l \neq k$ $A' \cap T_l \geq \min\{|X'_h \cap T_l|, \bar{q}_h^l\}$ or there is no such k.

□

Proof. This is an immediate consequence of proposition 11 and theorems 7 and 10. □

3.6 Distributional Constraints in Residency Matching

Residency matching is among the most important success stories of deferred acceptance. Prior to its implementation, the market for medical residents was badly unraveled Roth (2008). It was not uncommon for medical students to be hired almost two years before they graduated. Furthermore, the employment offers sometimes came in the form of "exploding offers" in which the resident had to respond immediately in the affirmative to assure a spot. Without such an immediate commitment, the offer would disappear. These were ostensibly designed to limit the information available to residents about other offers they might receive. In response to these issues, the National Residency Matching Program (NRMP) was developed to facilitate the match using a centralized clearinghouse. Through experimentation, they independently discovered the DA algorithm.

However, despite the success of DA in residency matching, it was observed that rural hospitals are often left with unfilled slots after the match Roth (1986). As residents make up a substantial share of the hospital labor force, this presents a problem for rural regions in treating their patients effectively. Applied matchmakers suggested that alterations to the algorithm might be helpful in ameliorating this problem. However, Roth (1986) showed that, in fact, if any hospital has vacant seats at the end of the deferred acceptance algorithm, they will have the same number of vacant seats at any stable match. Furthermore, the doctors matched to any such hospital will be the same at any stable match.

Motivated by this problem, the Japanese government developed a new mechanism for matching doctors. They imposed "regional caps" on each prefecture of Japan. The regional cap is an upper limit on the number of doctors that can be matched with each prefecture[20] (Kamada and Kojima 2015). Hospitals within each region were then assigned quotas such that the sum of hospital quotas within each region did not exceed the cap. The idea was to cap the enrollment of urban prefectures, thereby increasing the number of new residents to be matched with rural prefectures. Kamada and Kojima (2015) convincingly demonstrate the drawbacks of this mechanism and develop an alternative mechanism. They allow the hospital caps to be assigned flexibly throughout the mechanism.

The model here is motivated by Kamada and Kojima (2019) which generalizes the that presented in Kamada and Kojima (2015).

Suppose $X = D \times H$ and there is a finite set R of regions which partitions the hospitals. Let H_r denote the set of hospitals in region r and let $n_r = |H_r|$ for each region. For each region, there is assumed to be a regional cap q_r given by the matchmaker. I will assume that each region is associated with a linear order $\succ_r$ over $\mathbb{Z}_+^{n_r}$ which specifies the regional preference over possible distributions of doctors to each hospital within the region. Given an element ω of $\mathbb{Z}_+^{n_r}$, let $C_r(\omega) = \max_{\succ_r}\{\omega' \in \mathbb{Z}_+^{n_r} : \omega' \leq \omega\}$. In words, $C_r(\cdot)$ treats the discrete polytope $\{\omega' \in \mathbb{Z}_+^{n_r} : \omega' \leq \omega\}$ as the choice set and chooses a distribution of doctors which maximizes its preferences in that set. The idea here is that, over the course of the standard deferred acceptance algorithm, the cumulative proposals of doctors serves as an opportunity set of the regions. They can choose any distribution of doctors which does not exceed the number of applications to any hospital. I will assume that $\succ_r$ has the following properties for each region[21]:

1. $\omega' \succ_r \omega$ if $\omega_h > q_h \geq \omega'_h$ for some $h \in H_r$ and $\omega'_{h'} = \omega_{h'}$ if $h \neq h'$
2. $\omega' \succ_r \omega$ if $\sum_{h \in H_r} \omega_h > q_r \geq \sum_{h \in H_r} \omega_h$

In words, the first condition requires that the regional preferences never prefer to exceed the physical capacity of any hospital. The second condition requires that the regional preferences respect the regional quotas.

[20] There are 47 prefectures which partition the country.

[21] These are conditions (1) and (2) from Kamada and Kojima (2019), page 11

If $X' \subset X$, let $\omega_r(X') = (|X'_h|)_{h \in H_r}$ so the function ω_r simply counts the number of doctors for each hospital in region r as specified by X' and arranges those counts into a vector. Consider the constraint correspondence given by

$$B_h(X') = \{Y \in \sigma_h(X') : |Y| \leq \tau_h C_r(\omega_r(X'))\}$$

where τ_h is the coordinate projection corresponding to h. This can be understood as follows: given a collection of contracts X', the constraint correspondence first counts the number of doctors assigned to each hospital, then uses the preferences of regions to determine the constrained optimal distribution of doctors within each region and finally allows each hospital to choose a collection of contracts which replicates that distribution.

Kamada and Kojima (2019) assume furthermore that for all regions r, the choice rule C_r satisfies the condition that for all $\omega, \omega' \in \mathbb{Z}_+^{n_r}$:

$$\omega \leq \omega' \implies C_r(\omega) \geq C_r(\omega') \wedge \omega$$

which intuitively states that as ω increases to ω', the choice to increase the seats available to h should only increase if h is already being assigned ω_h.[22] I will call this the *KK-substitutes condition*. The following proposition shows that this condition is sufficient to ensure that B, as defined above, satisfies the generalized substitutes condition.

Proposition 13. *If $\succ_r$ satisfies the KK-substitutes condition, B, as defined above satisfies generalized substitutes.*

Proof. Let (A', X', A'', X'') be a monotone choice pair for h with $y \in A'' \setminus A'$ and $y \in X'$. Since h's option set is only limited in terms of the number of contracts that h can sign, it is sufficient to show that $B_h(X')$ is nonempty. Since in this case, either $d(y) \in d(A)$, and the associated contract in A can be swapped out for y or $d(y) \notin A$ and y can be swapped out for any contract in A. However, since $y \in A''$, $B_h(X'')$ is nonempty. Hence $\tau_h C_r(\omega_r(X''))\}$ is greater than or equal to one. Either $\tau_h C_r(\omega_r(X')) = \tau_h C_r(\omega_r(X''))$ or $\tau_h C_r(\omega_r(X')) < \tau_h C_r(\omega_r(X''))$ and in the first case, I get the result immediately. In the second, it follows by KK-substitutes. □

[22] See Kamada and Kojima (2019) for a detailed discussion on this condition.

Bibliography

Abdulkadiroğlu, A. (2005): "College admissions with affirmative action," *International Journal of Game Theory*, 33(4), 535–549.

Abdulkadiroglu, A. (2013): "School choice," *The Handbook of Market Design*.

Abdulkadiroglu, A., and T. Sönmez (2003): "School choice: A mechanism design approach," *The American Economic Review*, 93(3), 729–747.

Abdulkadiroğlu, A., P. Pathak, A. E. Roth, and T. Sonmez (2006): "Changing the Boston school choice mechanism," Discussion paper, National Bureau of Economic Research.

Abdulkadiroğlu, A., and T. Sönmez (1999): "House Allocation with Existing Tenants," *Journal of Economic Theory*, 88, 233–260.

Abraham, D. J., and D. F. Manlove (2004): "Pareto Optimality in the Roommates Problem," Discussion paper, Department of Computing Science, University of Glasgow.

Balbuzanov, I. (2019): "Constrained Random Matching," Working paper, University of Melbourne.

Barberà, S. (1983): "Strategy-Proofness and Pivotal Voters: A Direct Proof of the Gibbard- Satterthwaite Theorem," *International Economic Review*, 24, 413–418.

——— (2001): "An Introduction to Strategy-proof Social Choice Functions," *Social Choice and Welfare*, 18, 619–653.

Barberà, S., D. Berga, and B. Moreno (2010): "Individual versus Group Strategy-Proofness: When Do They Coincide?," *Journal of Economic Theory*, 145, 1648–1674.

——— (2016): "Group Strategy-Proofness in Private Good Economies," *American Economic Review*, 106, 1071–1099.

Bird, C. G. (1984): "Group Incentive Compatibility in a Market with Indivisible Goods," *Economics Letters*, 14(4), 309–313.

Bogomolnaia, A., and H. Moulin (1990): "A New Solution to the Random Assignment Problem," *Journal of Economic Theory*, 100, 295–328.

Budish, E., Y.-K. Che, F. Kojima, and P. Milgrom (2013): "Designing Random Allocation Mechanisms: Theory and application," *American Economic Review*, 103, 585–623.

Ehlers, L., I. E. Hafalir, M. B. Yenmez, and M. A. Yildirim (2013): "School Choice with Controlled Choice Constraints: Hard Bounds versus Soft Bounds," *Journal of Economic Theory*, 153, 648–683.

Erdil, A., and T. Kumano (2019): "Efficiency and stability under substitutable priorities with ties," *Journal of Economic Theory*, p. 104950.

Gale, D., and L. Shapley (1962): "College Admissions and the Stability of Marriage," *American Mathematical Monthly*, 69, 9–14.

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

Hafalir, I. E., M. B. Yenmez, and M. A. Yildirim (2013): "Effective Affirmative Action in School Choice," *Theoretical Economics*, 8, 325–363.

Hatfield, J. W., and P. R. Milgrom (2005): "Matching with Contracts," *American Economic Review*, 95(4), 913–935.

Irving, R. W. (1985): "An Efficient Algorithm for the 'Stable Roommates' Problem," *Journal of Algorithms*, 6, 577–595,.

Kamada, Y., and F. Kojima (2015): "Efficient Matching under Distributional Constraints: Theory and Applications," *American Economic Review*, 105, 67–99.

——— (2017a): "Recent Developments in Matching with Constraints," *American Economic Review Papers and Proceedings*, 107, 200–204.

——— (2017b): "Stability Concepts in Matching under Distributional Constraints," *Journal of Economic Theory*, 168, 107–142.

——— (2018): "Stability and Strategy-Proofness for Matching with Constraints: A Necessary and Sufficient Condition," *Theoretical Economics*, 13, 1761–794.

Kamada, Y., and F. Kojima (2019): "Accommodating various policy goals in matching with constraints," *The Japanese Economic Review*, pp. 1–33.

Kojima, F. (2012): "School choice: Impossibilities for affirmative action," *Games and Economic Behavior*, 75(2), 685–693.

Le Breton, M., and V. Zaporozhets (2009): "On the Equivalence of Coalitional and Individual Strategy-Proofness Properties," *Social Choice and Welfare*, 33, 287–309.

Maskin, E. (1999): "Nash Equilibrium and Welfare Optimality," *Review of Economic Studies*, 66, 23–38.

Meng, D. (2019): "Dictatorship and Connectedness for Two-Agent Mechanisms with Weak Preferences," Working paper, Southwest Baptist University.

Muller, E., and M. Satterthwaite (1977): "The Equivalence of Strong Positive Association and Strategy-proofness," *Journal of Economic Theory*, 14, 412–418.

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

Pathak, P. A., and T. Sönmez (2008): "Leveling the playing field: Sincere and sophisticated players in the Boston mechanism," *The American Economic Review*, 98(4), 1636–1652.

Pycia, M., and U. Ünver (2017): "Incentive Compatible Allocation and Exchange of Discrete Resources," *Theoretical Economics*, 12, 287–329.

Roth, A. E. (1985): "The college admissions problem is not equivalent to the marriage problem," *Journal of economic Theory*, 36(2), 277–288.

——— (1986): "On the allocation of residents to rural hospitals: a general property of two-sided matching markets," *Econometrica: Journal of the Econometric Society*, pp. 425–427.

——— (2002): "The economist as engineer: Game theory, experimentation, and computation as tools for design economics," *Econometrica*, 70(4), 1341–1378.

——— (2008): "Deferred acceptance algorithms: History, theory, practice, and open questions," *international Journal of game Theory*, 36(3-4), 537–569.

Roth, A. E., and M. Sotomayor (1992): "Two-sided matching," *Handbook of game theory with economic applications*, 1, 485–541.

Rothkopf, M. H. (2007): "Thirteen reasons why the Vickrey-Clarke-Groves process is not practical," *Operations Research*, 55(2), 191–197.

Satterthwaite, M. (1975): "Strategy-proofness and Arrow's conditions: Existence and Correspondence Theorems for Voting Procedures and Social Welfare Functions," *Journal of Economic Theory*, 10, 187–217.

Satterthwaite, M. A., and H. Sonnenschein (1981): "Strategy-proof allocation mechanisms at differentiable points," *The Review of Economic Studies*, 48(4), 587–597.

Shapley, L., and H. Scarf (1974): "On Cores and Indivisibility," *Journal of Mathematical Economics*, 1, 23–37.

Svensson, L.-G. (1994): "Queue Allocation of Indivisible Goods," *Social Choice and Welfare*, 11, 323–330.

——— (1999): "Strategy-Proof Allocation of Indivisible Goods," *Social Choice and Welfare*, 16, 557–567.

Takagi, S., and S. Serizawa (2010): "An impossibility theorem for matching problems," *Social Choice and Welfare*, 35(2), 245–266.

Takamiya, K. (2001): "Coalition Strategy-Proofness and Monotonicity in Shapley-Scarf Housing Markets," *International Journal of Game Theory*, 41, 115–130.

——— (2003): "On Strategy-Proofness and Essentially Single-Valued Cores: A Converse Result," *Social Choice and Welfare*, 20, 77–83.

——— (2013): "Coalitional Unanimity versus Strategy-proofness in Coalition Formation Problems," *Mathematical Social Sciences*, 42, 201–213.

Toda, M. (2006): "Monotonicity and Consistency in Matching Markets," *International Journal of Game Theory*, 34, 13–31.