

# Incentives and Efficiency in Constrained Allocation Mechanisms\*

Joseph Root<sup>†</sup>

David S. Ahn<sup>‡</sup>

Draft Updated January 9, 2020  
[Click Here For Most Recent Version](#)

## Abstract

We study private-good allocation mechanisms where an arbitrary constraint delimits the set of feasible joint allocations. This 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 for every possible constraint, showing that every mechanism is a “local dictatorship.” For more than two agents, we leverage this result to provide a 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 strategy-proof and Pareto efficient. We apply these results to the roommates problem to generate 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 new proof of the Gibbard–Satterthwaite Theorem. Finally, we introduce and study a large class of “constraint-traversing” mechanisms which can be defined for any constraint and we provide a simple sufficient condition for such mechanisms to be group strategy-proof and Pareto efficient. We construct constraint-traversing mechanisms for a number of examples.

## 1 Introduction

Every market is situated in a unique social, legal, and technological environment where both individual consumption choices and broader social allocations are subject to constraints. Increasingly, mechanism and market designers are working to expand the practical applicability of existing models by taking these constraints seriously (Roth 2002). In school choice, mechanisms are designed to take diversity considerations into account; in medical residency markets that match new doctors to hospitals, some governments require that assignments satisfy distributional requirements to guarantee that there are enough doctors for each region; in spectrum auctions, the allocation of radio frequency is designed to satisfy a large number of engineering conditions to ensure minimal cross-channel interference. While

---

\*We thank Haluk Ergin, Yuichiro Kamada, Chris Shannon, and seminar participants at Berkeley, Duke, McGill, and Yale for helpful feedback.

<sup>†</sup>Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: jroot@econ.berkeley.edu

<sup>‡</sup>Department of Economics, University of California, Berkeley, 530 Evans Hall, Berkeley, CA 94720-3880. Email: dahn@econ.berkeley.edu

specific approaches have been successfully tailored for each of these particular problems, little progress has been made developing a general or systematic approach to studying constraints in mechanism design. A comprehensive approach is important not only because it will enable analytical insights to be shared between contexts, but also because, in many real-world settings, the unpredictable variety of reasonable constraints that might be encountered curbs the usefulness of mechanisms designed for any specific constraint. Consider a firm assigning projects to workers where each project is required to be matched with a specific composition of worker types. For example, a consulting firm might need to assign a certain number of lawyers, experts and managers to each account. However, complementarities between different types of workers might relax the constraint. A lawyer with expertise on a given account might be able to replace both a lawyer and an expert. A high-level manager could substitute for multiple lower-level managers. Teams with experience working together might generate efficiencies that require fewer workers per account. In this context, the market designer has little a priori knowledge of the types of constraints to prepare for, because the contextual details are paramount. Instead, they need to be equipped with mechanisms designed for a wide variety of constraints in order to provide a satisfactory solution.

To that end, we develop a model of object allocation under completely general constraints. In our model, a finite number of objects are to be allocated to a finite number of agents and an arbitrary constraint circumscribes the set of feasible social allocations. Each agent is assumed to have strict preferences over the objects assigned to her, but is indifferent between any two social allocations in which she gets the same object. While another agent’s consumption imposes no direct effect on one’s well-being, her allocation does limit the profiles of allocations that are jointly feasible. So the constraint captures the indirect interactions across agents’ allocations. For each constraint, our goal is to study the set of incentive compatible and efficient mechanisms which yield feasible outcomes, regardless of agents’ preferences. In addition, and somewhat more abstractly, we study the properties of the correspondence which maps constraints to the set of desirable mechanisms, for example providing information about which constraints will yield a variety of mechanisms and which will not.

In addition to the practical advantages of this general approach to constraints, several prominent problems which at first glance appear unconstrained and unrelated to each other can actually be viewed as special constraints in our setting. For example, the classical social choice problem corresponds to the special constraint where all agents are required to get the same object.<sup>1</sup> From this perspective, the social choice problem may be seen as a special constrained private-goods allocation problem. In fact, a corollary of our results is the Gibbard–Satterthwaite Theorem which demonstrates that all strategy-proof social choice mechanisms are dictatorial. Having conceptualized the problem this way, we are also able to formulate and provide a converse to Gibbard–Satterthwaite: under what conditions does the constraint admit non-dictatorial mechanisms? Perhaps the second most studied special case of our model is the house allocation problem where a finite number of indivisible objects are to be allocated, with the constraint that each object be assigned to at most a single agent. 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<sup>2</sup> and Pareto efficient house allocation mechanisms, building on earlier

---

<sup>1</sup>Clearly, the term “object” is figurative. In social choice, the objects are usually policy choices or political candidates.

<sup>2</sup>Roughly, a mechanism is group strategy-proof if no coalition of agents can jointly misreport their preferences, without

work by Papái (2000). We briefly translate their results to our setting and show how to derive their characterization for a small number of agents. Another prominent model that can be expressed as a constraint is the roommates problem in which an even number of agents need to be matched into pairs. In this case, the “objects” are, in fact, the other agents, and the constraint requires both that no agent is matched to herself, and if  $i$  is matched to  $j$  then of course  $j$  is required to be matched with  $i$ . In contrast to the previous two examples, to our knowledge, no general characterization of the incentive compatible, efficient mechanisms has been developed in this setting. As an application of our results, we provide such a characterization. We show that a roommates mechanism is group strategy-proof and Pareto efficient if and only if it is a “generalized serial dictatorship,” which is a class of mechanisms we will formally introduce later.<sup>3</sup> The fact that so many well-known problems present themselves as constraints in our model is a fortunate and, we must admit, initially unanticipated, 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–Satterthwaite 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 an alternate reconciliation of these positive results by interpreting these environments as relaxing constraints in the *image* of the mechanism: outside of the Arrowian 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.

Despite the generality we allow for the structure of the constraint, we are able to generate a full characterization of the mechanisms that satisfy standard incentive and efficiency desiderata. We start by limiting the analysis to just two agents. In this case, we find a surprisingly parsimonious characterization of the set of individually strategy-proof and Pareto efficient mechanisms for every possible constraint. Furthermore, one can exploit simple properties of the constraint to provide this characterization. We show that for every constraint, 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 to an agent, the 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 forced to choose an object compatible with the dictator’s top object. Of course, not all partitions will yield desirable mechanisms. The set of possible local dictatorship

---

harming anyone in the group and making at least one agent strictly better off.

<sup>3</sup>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.

assignments depends crucially on the structure of the constraint. We show that every constraint, after permuting the order of objects, can have its infeasible allocations “block diagonalized” to yield an immediate characterization. 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 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 for all constraints. Indeed, in the house allocation setting, the collection of all such mechanisms is still unknown. Nevertheless, if we restrict attention to the set of group strategy-proof and Pareto efficient mechanisms for each constraint, we can leverage the two-agent results to get a new characterization. Group strategy-proofness requires that no group of agents can ever collectively misreport their preferences, making all agents in the group weakly better off, and at least one agent strictly better off. 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.” Surprisingly, the properties of the two agent marginal mechanisms are enough to capture the group incentives of a mechanism. In particular, if all two agent marginal mechanisms are Pareto efficient and individually strategy-proof, the full mechanism is group strategy-proof. This is especially useful given our explicit characterization of two-agent mechanisms with these properties. We therefore show that the two-agent mechanisms form the “building blocks” of all group strategy-proof mechanisms.

Beyond its analytical appeal, group strategy-proofness is natural and appealing 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.” In bossy mechanisms, agents can change the outcome of other agents without affecting their own allocation. Therefore, the marginal effect of restricting attention to group strategy-proofness, relative to requiring only individual strategy-proofness, is simply to rule out mechanisms with this property. 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 highly appealing properties, has largely not been put into practice, in part because of its susceptibility to group manipulation (Rothkopf 2007). We therefore believe that mechanisms with strong incentive properties are especially useful for practical considerations. At the least, group strategy-proofness is among the most demanding incentive conditions in the literature, and this benchmark must be established to understand the gains to efficiency from demanding weaker incentive conditions like Bayesian implementation. Finally, group strategy-proofness has been studied in other environments, and especially for the house allocation problem, so using this as our incentive condition will facilitate comparisons with earlier results. Nevertheless, our focus on Pareto-efficiency and group strategy-proofness rules out some practical mechanisms. Deferred acceptance, for example, is not

Pareto efficient, and is only individually strategy-proof for the proposing side. Even on the proposing side, it is not group strategy-proof.

While the characterization using marginal mechanisms is evidently useful for the problems described above, it can be practically difficult to construct new mechanisms for an arbitrary constraint, only knowing properties that the marginal mechanisms must satisfy. In fact, a large part of the literature in market design tries to find specific mechanisms that satisfy certain properties rather than to classify all such mechanisms. In this tradition, we conclude the paper by considering a large subclass of mechanisms, that we dub “constraint-traversing.” These mechanisms have the appealing property that they can be defined for any constraint, and often yield interesting mechanisms. A constraint-traversing mechanism is defined by a “local compromiser assignment,” which assigns a subset of agents to each infeasible allocation. For each preference profile, the mechanism uses the “constraint-traversing algorithm” to determine the ultimate allocation. This algorithm greedily attempts to give each agent their top choice and whenever it lands on an infeasible allocation it asks each agent from the set of local compromisers to choose their next-best alternative. The algorithm continues in this way until landing on a feasible allocation. Of course, not all local compromiser assignments will yield mechanisms with desirable properties. We therefore provide two conditions – forward and backward consistency – which, if satisfied, guarantee that the induced mechanism will be group strategy-proof and Pareto efficient. While we depart from the goal of full characterization in this section, the proof that forward and backward consistency are sufficient relies crucially on the results from the rest of the paper. We apply these mechanisms to a number of examples to demonstrate its utility for practical mechanism design problems.

## 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 example is the two-sided matching problem with distributional constraints, for example, where 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.<sup>4</sup> 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.<sup>5</sup> 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,

---

<sup>4</sup>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  $i$ .

<sup>5</sup>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).

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 until Pycia and Ünver (2017) recently provided an impressive full description of all group strategy-proof and Pareto efficient 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 Ünver. 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.<sup>6</sup> 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 Arrowian 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 Arrowian 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

---

<sup>6</sup>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.

almost the opposite of the house allocation constraint. We derive the Gibbard-Satterthwaite 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 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. Their 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. In fact, the difference in objectives is the most important part of our contribution. Barberà, Berga, and Moreno (2016) are primarily concerned with the relationship between group and individual incentives. Their main result uncovers a surprising connection between group and individual strategy-proofness when the space of admissible preferences is sufficiently rich. This complements earlier findings that illuminate a related connection for classic Arrovian environments, discovered by the same authors (Barberà, Berga, and Moreno 2010) and by Le Breton and Zaporozhets (2009). In contrast, our aim not to connect different axioms for strategy-proofness with each other, but rather to describe 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 expands if constraints are relaxed. 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, we mention a more distant body of work on constraints for random allocation that

examines when a random allocation is a convex combination of deterministic allocations satisfying certain constraints (Balbuzanov 2019, Budish, Che, Kojima, and Milgrom 2013), which would extend the fairness gains of the random assignment mechanisms introduced by Bogomolnaia and Moulin (1990). We focus on deterministic mechanisms, and as far as we can see our results have no relationship to these findings.

## 2 Model

Let us begin with some primitives.  $N$  is a finite set of **agents** and  $\mathcal{O}$  is a finite set of **objects**. We use the term “object” because of our leading examples, but note that  $\mathcal{O}$  are not necessarily physical objects, but can also be political candidates, roommates, etc. Define  $\mathcal{A} = \mathcal{O}^N$  to be the set of all possible allocations of objects to agents. We may equivalently think of  $\mathcal{A}$  as the set of maps  $\mu : N \rightarrow \mathcal{O}$  and we will switch between these equivalent perspectives as needed. A **suballocation** is a map  $\sigma : M \rightarrow \mathcal{O}$  where  $M \subset N$ . We will 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 ex ante. 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.<sup>7</sup> 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  $\mathcal{P} = P^N$  to denote the set of preference profiles.<sup>8</sup> Our primary object of interest in this paper is a **feasible mechanism**, which is simply a map  $f : \mathcal{P} \rightarrow C$ . 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

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

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).

- **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  $\mathcal{O} = N$

---

<sup>7</sup>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  $(a_i)_{i \in N} \in C$

<sup>8</sup>A binary relation  $B \subset \mathcal{O} \times \mathcal{O}$  is a linear order if it is complete, transitive, and antisymmetric



and imposing the constraint

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

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<sup>9</sup>. Specifically, if

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

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. This paragraph may be skipped and referred to as needed. For disjoint sets of objects  $A_1, A_2 \dots A_m$  we will denote

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

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  $P^\uparrow[\{a\}]$ . For any subset  $M \subset N$ , given a preference profile  $\succsim = (\succsim_i)_{i \in N} \in \mathcal{P}$  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  $C^M = \{\mu : M \rightarrow \mathcal{O} \mid \exists b \in C \text{ s.t. } b_i = \mu(i) \forall i \in M\}$  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  $\mu : M \rightarrow \mathcal{O}$  and  $\mu' : M' \rightarrow \mathcal{O}$  are suballocations with  $M \subset M'$  which agree on their shared domain,  $\mu'$  is called a **extension** of  $\mu$ . If  $\mu'$  is a feasible suballocation (which of course implies that  $\mu$  is) then  $\mu'$  is called a **feasible extension** of  $\mu$ . If  $\mu'$  assigns an object to each agent, it is called a **complete extension** of  $\mu$ . Given a feasible suballocation  $\mu$ , we will let  $C(\mu)$  denote the set of complete and feasible extensions of  $\mu$ . For any agent  $i$ , let  $\pi_i : \mathcal{A} \rightarrow \mathcal{O}$  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  $\pi_i B = \{a \in \mathcal{O} \mid \text{there is a } b \in B \text{ with } \pi_i b = a\}$ . For  $x \in \mathcal{O}$  and  $\succsim_i \in P$ , define  $LC_{\succsim_i}(x) = \{y \in \mathcal{O} \mid y \prec_i x\}$  be the (strict) **lower contour set** of  $x$  at  $\succsim_i$ . Likewise,  $UC_{\succsim_i}(x) = \{y \in \mathcal{O} \mid y \succ_i x\}$  is the (strict) **upper contour set** of  $x$  at  $\succsim_i$ . For a preference  $\succsim_i$ , define  $\tau_n(\succsim_i)$  as the  $n$ th top choice under  $\succsim_i$ . Likewise, for any preference profile  $\succsim$ ,

<sup>9</sup>See Barberà (2001) for a general statement of the social choice problem with restricted domains.

define  $\tau_n(\succsim)$  as the allocation in which each agent gets their  $n$ th top choice. To save on notation, we will often omit the subscript when referring to 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 ensure that some additional efficiency and incentive properties are satisfied. Below we list a number of well-known and potentially desirable features of allocation mechanisms.

**Definition 1.** A mechanism  $f : \mathcal{P} \rightarrow C$  is

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

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

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

2. **group strategy-proof** if, for every  $\succsim \in \mathcal{P}$  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  $\succsim \in \mathcal{P}$  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  $a_j \succ_j f(\succsim)$  for all  $j$ .

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

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

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

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

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 “unanimity” 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  $f : \mathcal{P} \rightarrow \mathcal{A}$  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.<sup>10</sup> 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–Satterthwaite Theorem, which proceeds by showing that either group or individual strategy-proofness is equivalent to Maskin monotonicity for the social choice problem (Muller and Satterthwaite 1977).<sup>11</sup> 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

<sup>10</sup>For the specific problem of house allocation, the equivalence between (1) and (2) was first observed by Papái (2000).

<sup>11</sup>Recall the Muller–Satterthwaite Theorem: all Maskin monotonic and surjective social choice functions are dictatorial.

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  $f : \mathcal{P} \rightarrow \mathcal{A}$  is group strategy-proof then it is Pareto efficient on its image.*<sup>12</sup>

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  $C$ ,  $GS(C)$ .

### 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.

#### 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 pick 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, 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  $R_1 = \{x \in \mathcal{O} \mid (x, y) \notin C \text{ for all } y \in \mathcal{O}\}$

<sup>12</sup>That is, if the constraint  $C$  is exactly  $im(f)$ .

and  $R_2 = \{y \in \mathcal{O} \mid (x, y) \notin C \text{ for all } x \in \mathcal{O}\}$ . 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  $f : \mathcal{P} \rightarrow C$  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  $\bar{C}^* = \{(x, y) \mid (x, y) \notin C \text{ and } x \notin R_1, y \notin R_2\}$ . 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  $f : \mathcal{P} \rightarrow C$  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  $B$ .<sup>13</sup> Since  $B$  is reflexive and symmetric, it can easily be shown that  $T$  is an equivalence relation.<sup>14</sup> As an equivalence relation on a finite set, it can be expressed as a partition with a finite number of equivalence classes  $E_1, E_2, \dots, E_p$ , 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 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 equivalence 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.

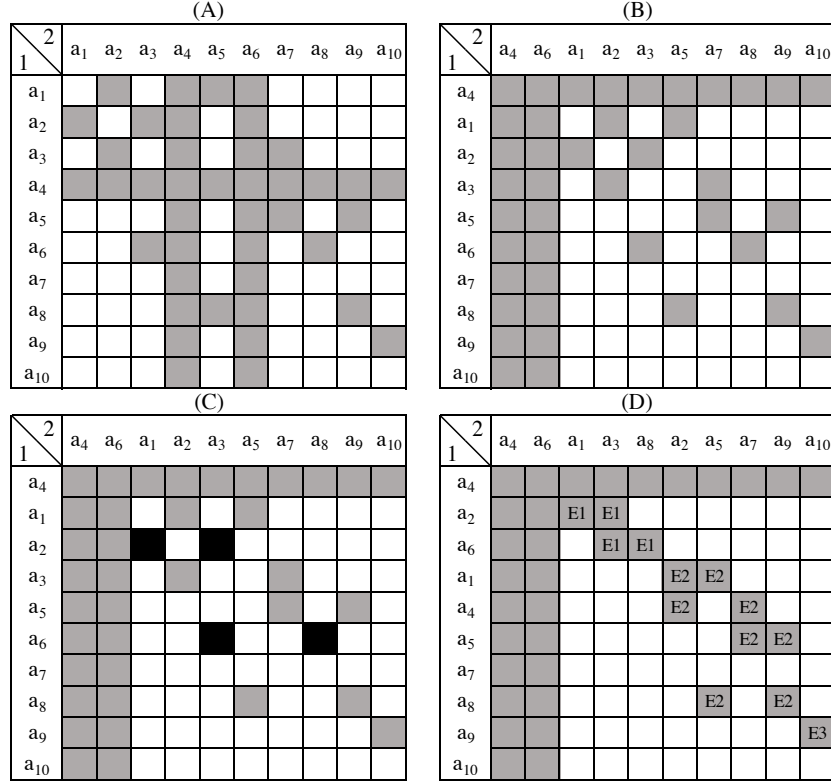


Figure 1: Two-agent Example

<sup>13</sup>The transitive closure is the minimum binary relation containing  $B$  which is transitive.

<sup>14</sup>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  $(a_1, b_1), \dots, (a_n, b_n)$  such that  $(a, b)B(a_1, b_1)B \dots B(a_n, b_n)B(a', b')$ . By reversing all these, we see that  $(a', b')T(a, b)$ .

A mechanism  $f : \mathcal{P}^2 \rightarrow C$  is called a **local dictatorship** if each block  $E_i$  of  $\bar{C}^*$  is assigned a (local) dictator  $d_i$  so that for any  $\succsim$  if  $\tau(\succsim_1, \succsim_2) = (a, b)$

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

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.**  $f : P^2 \rightarrow C$  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 2. On the left is the house allocation constraint and on the right is the social choice constraint. It's clear that each grey square on the left is a different equivalence class of  $T$ , so every mechanisms 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 gives 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 another perspective on this requirement: observe that if the social choice constraint in Figure 2 had only two objects, the constraint would be the top-left  $2 \times 2$  constraint. In this case,  $T$  now has two equivalence classes corresponding to the two grey squares.

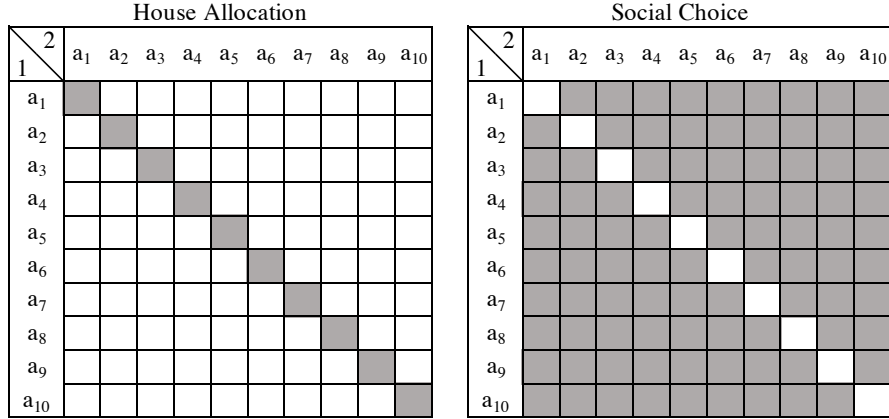


Figure 2: The social choice and house allocation constraints for two agents and 10 objects.

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 one of the two agents as 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 indifferences and ours for constraints, so their substantive applications and contributions are quite different.

### 3.2 N Agents

When there are three or more agents, the approach we used for two agents fails to provide a straightforward characterization and we therefore take a somewhat less direct approach. However, there is a subclass of constraints for which the characterization is no more difficult than in the two-agent case. A constraint  $C$  is called **single-compromising** if for every infeasible  $(a_i)_{i \in N}$ , for every  $i$  there is a  $a'_i$  such that  $(a'_i, a_{-i})$  is feasible. Thus, from any infeasible allocation, all agents have the possibility of unilaterally compromising 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 colors 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

$$\alpha : \mathcal{A} \rightarrow 2^N$$

such that for every  $x \in \bar{C}$ ,  $\alpha(x)$  is nonempty and for every  $y \in C$ ,  $\alpha(y) = \emptyset$ . 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  $\alpha$ , 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  $\alpha$ . 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 now, we will ignore the complication that the mechanism could potentially exhaust an agent's objects. We will delay a more



detailed discussion until Section 5, where we will introduce a sufficient condition on the local compromiser assignment that guarantees this complication does not arise. But since this complication is not germane when the constraint is single-compromising, we sidestep this problem for the moment and simply say that 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  $(a'_i, a_{-i}) \in \bar{C}$

As mentioned before, we table a thorough treatment of constraint-traversing mechanisms until Section 5.

From hereon, we are considering 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  $f : \mathcal{P} \rightarrow C$  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  $f_{\succsim_{M^c}}^M : P^M \rightarrow \mathcal{O}^M$  and is defined by

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

we will denote  $I^M(\succsim_{M^c}) = \text{im}(f_{\succsim_{M^c}}^M)$  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  $f_{\succsim_{M^c}}^M(\succsim) \in \mathcal{O}^M$ .

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  $f : \mathcal{P} \rightarrow C$  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.

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. Let  $F_n = \{f : P^n \rightarrow \mathcal{O}^n\}$  and  $\mathcal{E}_{n,m} = \{\phi : P^n \rightarrow GS(\mathcal{O}^m)\}$ . 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  $\sigma \in \mathcal{E}_{n,m}$  provides, for each preference profile of  $n$  agents, a group strategy-proof mechanism for  $m$  other agents. Likewise, define  $\mathcal{F}_{n,m} = \{\eta : P^n \rightarrow F_m\}$ . We will need the following definition:

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

$$f \oplus g(\succ) = [f(\succ_1, \succ_2, \dots, \succ_n), g(\succ_{n+1}, \succ_{n+2}, \dots, \succ_{n+m})]$$

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

$$\succ \mapsto [\rho(\succ_{n+1}, \dots, \succ_{n+m})(\succ_1, \dots, \succ_n), \sigma(\succ_1, \dots, \succ_n)(\succ_{n+1}, \dots, \succ_{n+m})]$$

The final claim records these observations, explicitly providing a formula that characterizes the set of group strategy-proof mechanisms.

**Corollary 1.**

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

Where  $\text{Sym}(N)$  is the set of permutations of the agents  $N$ .

## 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.

### 4.1 Generalized Serial Dictatorship

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

**Definition 4.** Let  $\sigma(1), \dots, \sigma(N)$  be a strict ordering of the agents  $\{1, 2, \dots, N\}$ . 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  $\mu_1$  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  $\mu_k$  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.<sup>15</sup> 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  $\mu : M \rightarrow \mathcal{O}$  where  $M \subset N$ ). Let  $\mathcal{S}'$  be the set of incomplete suballocations<sup>16</sup>. A **GSD-ordering** is a map  $\tau : \mathcal{S}' \rightarrow N$  such that for any suballocation  $\mu$ ,  $\tau(\mu)$  is an agent not allocated an object under  $\mu$ . 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  $d_1 \equiv \tau(\emptyset)$  is the first dictator. She chooses her favorite object  $a_1$  from  $\pi_{d_1}C$ . Let  $\mu_1$  be the suballocation in which  $d_1$  is assigned  $a_1$  and all other agents are unassigned.
- Step  $k$**  The agent  $d_k \equiv \tau(\mu_{k-1})$  chooses her favorite object  $a_k$  from  $\pi_{d_k}C(\mu_{k-1})$ . Let  $\mu_k$  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  $\tau(\emptyset) = \sigma(1)$ ,  $\tau(\mu) = \sigma(2)$  for all suballocations  $\mu$  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. This is because the GSD-ordering  $\tau$  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.

<sup>15</sup>A fact we will prove shortly.

<sup>16</sup> $M$  is a proper subset of  $N$ .

**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.<sup>17</sup>

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  $\tau$ . 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  $(f^M, \tau) : \mathcal{P} \rightarrow C$  via the following algorithm:

**Step 1** Allocate  $f_i^M(\succsim_M)$  to every agent  $i$  in  $M$ . Let  $\mu_0$  this suballocation. Let agent  $d_1 = \tau(\mu_0)$  choose her favorite object  $a_1$  from among  $\pi_{d_1} C(\mu_0)$  and let  $\mu_1$  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  $d_k \equiv \tau(\mu_{k-1})$  chooses her favorite object  $x_k$  from  $\pi_{d_k} C(\mu_{k-1})$ . Let  $\mu_k$  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  $f^M : P^M \rightarrow C^M$  is Pareto efficient and group strategy-proof, for any GSD-ordering  $\tau$ , the mechanism  $(f^M, \tau)$  is group strategy-proof and Pareto efficient.*

## 4.2 The Roommates Problem

Suppose that there are an even number of agents who need to be paired up into roommates. Each agent has a strict preference over their possible matches. As discussed above, we can model this in our environment by letting  $\mathcal{O} = N$  and using the constraint

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

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, 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.

<sup>17</sup>So long as the constraint is nonempty, which we assume throughout.

**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. Our constructive approach shows how this tension is related to the structure of the roommates problem as a constraint in our more general environment.

### 4.3 Social Choice

The first observation in implementation theory was the celebrated negative result of Gibbard (1973) and Satterthwaite (1975) that, for the social choice problem with a full domain of preferences, the only strategy-proof and surjective 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 Gibbard–Satterthwaite Theorem yields additional insights that help understand the theorem more deeply. First, our approach to the theorem, in an environment 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–Satterthwaite 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.<sup>18</sup>

---

<sup>18</sup>This observation can also be alternatively deduced directly from the Gibbard–Satterthwaite Theorem, since dictatorships are both individual and group strategy-proof. Since our aim is to prove that theorem, this is clearly not valid

**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  $f : \mathcal{P} \rightarrow C$  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  $f : \mathcal{P} \rightarrow C$  is surjective and strategy-proof then it is dictatorial.*<sup>19</sup>

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 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.*

## 5 Constraint-Traversing Mechanisms

We introduced the notion of a constraint-traversing mechanism in Section 3. 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. 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  $\alpha$  is such that the constraint-traversing algorithm terminates in an allocation for any preference profile, we say  $\alpha$  is **implementable**. An implementable  $\alpha$  induces a mechanism  $f^\alpha$ , in which every preference profile yields the allocation derived from the associated constraint-traversing algorithm. Conversely, a **constraint-traversing mechanism** is a feasible mechanism  $f : \mathcal{P} \rightarrow C$  such that there is some local compromiser assignment  $\alpha$  which induces it.

---

for our approach.

<sup>19</sup>In fact, we only need that  $|\text{im}(f)| > 2$  in which case we could drop items never allowed and recover the same statement.

An initial difficulty is that the local constraint assignment may not be unique. In Figure 3, 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.

(I)				(II)				(III)			
1 \ 2	a	b	c	1 \ 2	a	b	c	1 \ 2	a	b	c
a	1&2	1	1	a	1	1	1	a	2	1	1
b	2			b	2			b	2		
c	2			c	2			c	2		

Figure 3: 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  $\alpha'$  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  $\alpha$  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  $x$ <sup>20</sup>. The local compromiser assignment is complete in panel 1 above and is not complete 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  $f : \mathcal{P} \rightarrow C$  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  $f$ .

<sup>20</sup>Since no matter where the local compromisers go, this agent will need to compromise.

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  $\alpha$  for the induced mechanism (provided  $\alpha$  is implementable), to be group strategy-proof.

**Definition 6.** Given a local compromiser assignment  $\alpha$ , 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)  $\{i \mid z^l \neq z^{l-1}\} = \{i_l\} \subset \alpha(z^{l-1})$  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  $\alpha$  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  $x \in \mathcal{O}$ ,  $(x, z_{-j}^0) \in \bar{C}$  and  $\alpha(x, z_{-j}^0) \cap \{i_1, \dots, i_{p-1}\}$  is nonempty.*

*then the induced mechanism  $f^\alpha$  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  $\alpha$  which satisfies forward and backward consistency is implementable if and only if there is no monotone path  $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$  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.

## 5.1 Examples

### 5.1.1 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
1	1&2	1	1
a			
b	2	3	
c	2		

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

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

Figure 4: A non-GSD mechanism

Consider Figure 4. 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 4 is a local compromiser assignment which determines a mechanism. Both forward and backward compatibility can be easily checked, giving the following lemma:

**Lemma 5.** *The mechanism introduced in Figure 4 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  $a$ <sup>21</sup>. 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.

### 5.1.2 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 5. 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 nontrivial 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

<sup>21</sup>The allocations  $(a, c, c)$ ,  $(c, a, b)$  and  $(c, c, a)$  are all feasible

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.

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

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

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

Figure 5: Constraint for House Allocation Problem

In Figure 6 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.

1	1	1		1	2			1		3
2	2			1	1&2	1			1	3
3		1			2	1		3	1	1&3

1&2	1	1		1 or 2	2			2		2
2	1 or 2			1	1&2	1			2	2
2		2			2	2		3	3	2&3

1	1	1		1	2			1		3
2	2			1	2	3			2	3
3		1			2	2		3	3	3

1&2	1	1		1	3			2		2
2	1			1	1&3	3			1	3
2		2			3	2		3	3	2&3

1&2	1	1		1 or 2	2			1 or 2		2
2	1 or 2			1	1&2	1			1 or 2	2
2		1 or 2			2	1 or 2		1	1	1&2

Figure 6: All group strategy-proof and Pareto efficient 3-Agent Mechanisms (up to symmetry)

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 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 7.

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

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

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

Figure 7: 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 8. We demonstrate the process for constructing all of them in the Supplemental Appendix.

	3	1		3	3			3		3
2 or 3	2			3	2&3	3			2	3
3		1			2	2		3	3	3

	1 or 3	1		1	3			1 or 2		2
1	1			1	1&3	1			1	2
1		1			3	1		1	1	1&2

	1	1		1 or 2	2			1 or 2		2
1	1			1	1&2	1			1 or 2	2
1		1			2	1 or 2		1	1	1&2

Figure 8: The Constraint-Traversing Mechanisms for the Constraint in Figure 7

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  $c$ . 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.

## 6 Conclusion

We studied the design of mechanisms to allocate a finite number of objects to a finite number of agents under an exogenous constraint. We first characterize the two-agent group strategy-proof and Pareto efficient mechanisms for any constraint. We then use that characterization to develop a new characterization for group strategy-proofness with an arbitrary number of agents. We apply this result to a number of applications including social choice, house allocation, and the roommates problem.

## A Appendix

### A.1 Proof of Proposition 1

We first need the following lemma, which is simply the forward direction of Barbera's observation:

**Lemma 6.** *Let  $f : \mathcal{P} \rightarrow \mathcal{A}$  be strategy-proof. Then for each  $i$  there is a map  $g_i : P^{n-1} \rightarrow 2^{\mathcal{O}} \setminus \{\emptyset\}$  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.  $\square$

*Proof.* (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  $\succsim, \succsim' \in \mathcal{P}$  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 6 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  $\succsim \in \mathcal{P}$  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

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

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  $\succsim_j^*$  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.  $\square$

## A.2 Proof of Lemma 1

By way of contradiction suppose that  $f : \mathcal{P} \rightarrow im(f)$  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  $\succsim$ .

## A.3 Proof of Lemma 2

Let  $\{g_i\}_{i \in N}$  be as in Lemma 6. 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.  $\square$

## A.4 Proof of Theorem 1 (Two-agent characterization)

( $\Leftarrow$ ) This direction follows by construction and by Barberà's Lemma.

( $\Rightarrow$ ) First note that Pareto efficiency implies that for all  $\succsim'_1$  and all  $\succsim'_2$ ,

$$g^1(\succsim'_2) \supset C^1(\max_{\succsim'_2} \mathcal{O}) \text{ and } g^2(\succsim'_1) \supset C^2(\max_{\succsim'_1} \mathcal{O})$$

since  $(\max_{\succsim_i} \mathcal{O})_{i=1,2} \in C$  implies  $f(\succsim) = (\max_{\succsim_i} \mathcal{O})_{i=1,2}$ .

If  $C = \mathcal{O}^2$  Pareto efficiency implies that  $f(\succsim) = (\max_{\succsim_1} \mathcal{O}, \max_{\succsim_2} \mathcal{O})$  for all  $\succsim$  which is (trivially) a local dictatorship. Suppose now that  $C$  is a proper subset of  $\mathcal{O}^2$  and let  $(a, b) \in \bar{C}^*$ . By assumption, there are  $a', b'$  such that  $(a, b')$  and  $(a', b)$  are in  $C$ . Let  $\succsim_1 \in \mathcal{P}^\uparrow[a, a']$  and  $\succsim_2 \in \mathcal{P}^\uparrow[b, b']$ . Since<sup>22</sup>  $im(f) = C$  from Barberà's Lemma we must have either  $a \notin g^1(\succsim_2)$  or  $b \notin g^2(\succsim_1)$ . Without loss, assume  $a \notin g^1(\succsim_2)$ . By efficiency we then have  $f(\succsim_1, \succsim_2) = (a', b)$ . By Maskin monotonicity, we have that  $f(\succsim_1, \succsim'_2) = (a', b)$  for any  $\succsim'_2 \in \mathcal{P}^\uparrow[b]$  hence  $a \notin g^1(\succsim'_2)$ . Therefore if  $C^1(b) = \mathcal{O} \setminus \{a\}$  we have that  $g^1(\succsim'_2) = C^1(b)$  for any  $\succsim'_2 \in \mathcal{P}^\uparrow[b]$  and efficiency implies that  $a \in g_2(\succsim'_1)$  for all  $\succsim'_1$ . Thus we may declare 2 the local dictator at  $(a, b)$ . If instead there is a distinct  $a'' \notin C^1(b)$ , let  $\succsim'_1 \in \mathcal{P}^\uparrow[a, a'', a']$ . Now we have that  $a \notin g^1(\succsim_2)$  so we must have either  $a'' \notin g^1(\succsim_2)$  or  $b \notin g^2(\succsim'_1)$ , however the outcome in the latter case would be Pareto-dominated by  $(a, b')$ . Hence  $a'' \notin g^1(\succsim_2)$ . Since  $a''$  was an arbitrarily chosen element of  $C^1(b)$  we have that  $g^1(\succsim_2) = C^1(b)$ . Efficiency then implies that  $b \in g^2(\succsim'_1)$  for all  $\succsim'_1$ . Finally, strategy-proofness gives that  $g^1(\succsim'_2) = C^1(b)$  for any other  $\succsim'_2 \in \mathcal{P}^\uparrow[b]$ . In this case we may declare 2 the local dictator at  $(a, b)$ .

We have proven that each incompatible pair  $(x, y)$  has a local dictator. It remains to prove that the dictator is constant within blocks. Consider  $(a, b)$  and  $(a, b')$  in  $\bar{C}^*$ . If 2 is the local dictator at  $(a, b)$  then whenever she puts  $b$  as her top choice she removes  $a$  from 1's option set. Hence 1 cannot be the dictator at  $(a, b')$ . If  $(a, b)$  and  $(a', b)$  are in  $\bar{C}^*$  then if 2 is a local dictator at  $(a, b)$  then when she puts  $b$  on top she removes  $a$  and  $a'$  from 1's option set. Thus 1 cannot be dictator at  $(a', b)$ . Since every point in the complement of the constraint must be assigned a dictator, we see that the dictatorship is constant along equivalence classes of  $T$ .  $\square$

<sup>22</sup>This follows immediately from the definition of Pareto efficiency.

## A.5 Proof of Proposition 2

First we show that every group strategy-proof, Pareto efficient mechanism is constraint-traversing. Let  $C$  be a single-compromising constraint and fix and a group strategy-proof, Pareto efficient mechanism  $f : \mathcal{P} \rightarrow C$ . 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 6 we have that for any  $\succsim'_{-k}$  with  $\max_{\succsim'_j} \mathcal{O} = a_j$  for all  $j \neq k$ ,  $a_k \notin g_k(\succsim'_{-k})$ <sup>23</sup>, 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  $\alpha$  on all of  $\bar{C}$ . Finally, we establish inductively that  $f$  is constraint-traversing according to  $\alpha$ . 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  $\alpha$  works, giving the result.

Now we need to show that  $\alpha$  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 the same 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  $(a'_k, a_{-k}) \in C$ . Then we would 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 a simple consequence of Maskin monotonicity and Proposition 1. □

## A.6 Proof of Theorem 2 ( $N$ -agent characterization)

If  $f$  is group strategy-proof, the marginal mechanisms are group strategy-proof by definition. To see 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  $f_{\succsim_{-i,j}}^{i,j}$  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  $f_i(\succsim'_i, \succsim_{-i}) = f_i(\succsim)$  then if 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  $f_j(\succsim'_i, \succsim_{-i}) \prec_i f_j(\succsim)$  in the former case  $i$  and  $j$  can announce  $(\succsim'_i, \succsim_j)$  when their true preferences are  $(\succsim_i, \succsim_j)$  and both weakly improve, with  $j$  strictly improving. Conversely, in the latter case they can report  $(\succsim_i, \succsim_j)$  when their true preferences are  $(\succsim'_i, \succsim_j)$  again to benefit  $j$ . □

<sup>23</sup>Recall the definition of the  $g_i$  from Lemma 6.

## A.7 Proof of Corollary 1

The proof is an immediate application of Theorem 2.  $\square$

## A.8 Proof of Proposition 3

Maskin monotonicity is easily seen to be satisfied. 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  $im(f) = C$ .  $\square$

## A.9 Proof of Proposition 4

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

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

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,  $(f^M, \tau)_i(\succsim') = (f^M, \tau)_i(\succsim)$  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  $(f^M, \tau)$  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  $(f^M, \tau) = (a_i)_{i \in N}$  as desired.  $\square$

## A.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 can ignore any agents’ ranking of themselves, which we will do. If  $N = 2$  there is only one possible allocation, so every mechanism is trivially a generalized serial dictatorship. Furthermore, if  $N = 4$  one can show that we can rewrite the problem as a social choice problem since a single agents’ 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.

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  $R_i = \{x \in N \mid (x, y) \notin I^{ij}(\succsim_{-ij}) \text{ for all } y \in N\}$  and  $R_j = \{y \in N \mid (x, y) \notin I^{ij}(\succsim_{-ij}) \text{ for all } x \in N\}$ . Then we get a marginal constraint like the one shown on the left of Figure 9



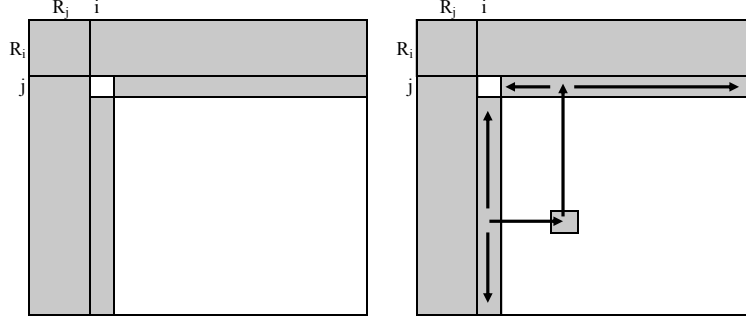


Figure 9:  $I^{ij}(\tilde{\succ}_{-ij})$

with the exception that some non-grey squares may actually be infeasible. 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 shown in on the right of Figure 9.<sup>24</sup> Therefore there must be a single dictator in the marginal mechanism  $f_{\tilde{\succ}_{-ij}}^{ij}$  by Theorems 1 and 2. Otherwise, every allocation in  $[N - R_i \cup \{j\}] \times [N - R_j \cup \{i\}]$  is infeasible or the set is empty. In the latter case, there is of course only one marginal mechanism. In the former case, there are three possible Pareto efficient, strategy-proof marginal mechanisms as illustrated in figure 10.

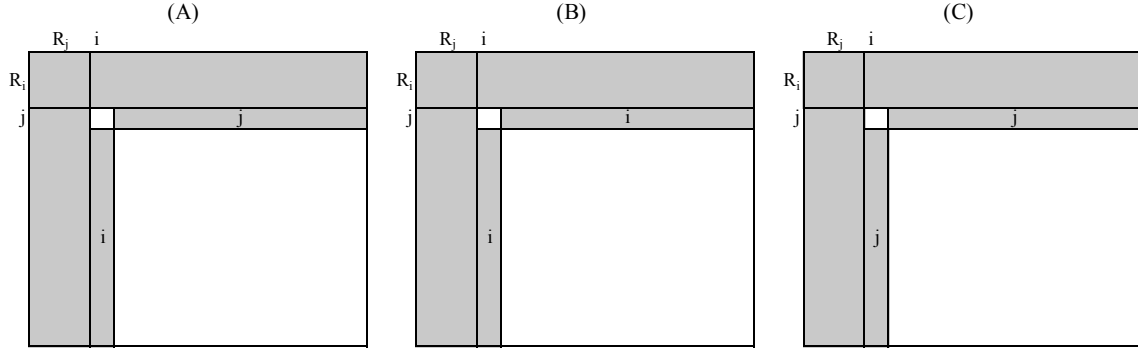


Figure 10: The three possible mechanisms  $f_{\tilde{\succ}_{-ij}}^{ij}$

Summarizing, if  $(j, i) \in I^{ij}(\tilde{\succ}_{-ij})$ , there are four possible types of mechanisms  $f_{\tilde{\succ}_{-ij}}^{ij}$ . (1) we could have  $\{(j, i)\} = I^{ij}(\tilde{\succ}_{-ij})$  so  $f_{\tilde{\succ}_{-ij}}^{ij}$  is constant ; (2)  $i$  could be the only dictator; (3)  $j$  could be the only dictator ;(4) we have the mechanism in panel (A) of figure 10.

We will now need the following lemma (which depends on the induction hypothesis).

**Lemma 7.** *Let  $A$  be a nonempty, proper subset of  $N$  with an even number of agents and  $|A| \geq 4$ . If  $\tilde{\succ}_{N \setminus A}^* \in [P^\uparrow(N \setminus A)]^{N \setminus A}$  then there is an agent  $j \in A$  such that*

$$f_j(\tilde{\succ}_A, \tilde{\succ}_{N \setminus A}^*) = \max_{\tilde{\succ}_j} N$$

whenever  $\tilde{\succ}_j N \in A$

<sup>24</sup>Recall the relation  $T$  was defined immediately before the statement of Theorem 1.

*Proof.* Suppose that we restrict attention to  $\succsim_A \in [P^\uparrow(A)]^A$ , then by Pareto efficiency, the agents in  $A$  are matched with one another, and the agents in  $N \setminus A$  are also matched to one another. But by group strategy-proofness and Pareto efficiency of  $f$ ,

$$f(\cdot, \succsim_{N \setminus A}^*)|_{[P^\uparrow(A)]^A}$$

gives a group strategy-proof and Pareto efficient roommates mechanism for the agents in  $A$ .<sup>25</sup> By the induction assumption, this mechanism is a generalized serial dictatorship. Thus there is a  $j$  such that  $f_j(\succsim_A, \succsim_{N \setminus A}^*) = \max_{\succsim_j} N$  whenever  $\succsim_A \in [P^\uparrow(A)]^A$ . Thus it remains to show, that  $j$  gets her top choice regardless of the announcements of the other agents in  $A$  so long as her top choice is in  $A$  and the agents in  $N \setminus A$  announce  $\succsim_{N \setminus A}^*$ . To see this, let  $\succsim_A$  be arbitrary except that  $\max_{\succsim_j} N \in A$ . For each  $i \neq j$  in  $A$ , let  $\succsim'_i \in P^\uparrow(A)$  put  $j$  top. Then  $g_j(\succsim_A, \succsim_{N \setminus A}^*) \supset A - \{j\}$ . Consider any  $i, j$  mechanism with  $i \in A$ . Since  $(i, j) \in I^{i,j}(\succsim'_i, \succsim_{N \setminus A}^*)$  there are four options for  $f_{(\succsim'_i, \succsim_{N \setminus A}^*)}^{ij}$ . However, the only one that fits with the fact that  $g_j(\succsim_A, \succsim_{N \setminus A}^*) \supset A - \{j\}$ , which given that  $|A| \geq 4$  leaves  $j$  the option of matching with agents other than  $i$  is that  $j$  is the only dictator in this marginal mechanism. Hence we have  $g_j(\succsim_i, \succsim'_i, \succsim_{N \setminus A}^*) \supset A - \{j\}$ . Repeating this argument one agent at a time gives that  $g_j(\succsim_A, \succsim_{N \setminus A}^*) \supset A - \{j\}$ , which is the desired result.  $\square$

Having done this, if  $A = \{3, 4, \dots, N\}$ ,  $\succsim_1^* \in P^\uparrow(2)$  and  $\succsim_2^* \in P^\uparrow(1)$ , there is a dictator among  $A$ . Without loss, suppose this dictator is 3 so that  $g_3(\succsim_1^*, \succsim_2^*, \succsim'_-) \supset A - \{3\}$  for all  $\succsim'_-$ . Now, for agents  $5, \dots, N$ , let  $\succsim_k^* \in P^\uparrow[\{5, \dots, N\}]$  then we also have a dictator among the  $\{1, 2, 3, 4\}$ . Now at the profile  $(\succsim_1^*, \succsim_2^*, \succsim_3, \succsim_4, \succsim_-)$ , where 3 top ranks 4, then 3 and 4 are matched regardless of 4's announcement. Hence the dictator can't be 4. This leaves three cases: the dictator could be 1, 2, or 3. Since, 1 and 2 are thus far symmetric, we will handle both of these cases at the same time.

First, however, we will start with the case where 3 is the dictator. In this case,  $g_3(\succsim'_1, \succsim'_2, \succsim'_4, \succsim_-^*) \supset \{1, 2, 4\}$  for all  $\succsim'_1, \succsim'_2$  and  $\succsim'_4$  by Lemma 7. In addition,  $g_3(\succsim_1^*, \succsim_2^*, \succsim'_4, \succsim_-^*) \supset N - \{1, 2, 3\}$ . Putting these together we have that  $g_3(\succsim_1^*, \succsim_2^*, \succsim'_4, \succsim_-^*) = N - \{3\}$ . Now, we know 3 is the marginal dictator in the  $(1, 3), (2, 3)$  and  $(3, 4)$  marginal mechanisms when all the other agents are announcing  $\succsim_-^*$ . Therefore,  $g_3(\succsim''_1, \succsim_2^*, \succsim'_4, \succsim_-^*) = N - \{3\}$  for any  $\succsim''_1$ . But again, 3 is the marginal dictator between 1 and 2, so  $g_3(\succsim''_1, \succsim''_2, \succsim'_4, \succsim_-^*) = N - \{3\}$  for any  $\succsim''_2$  and so on. Hence we actually have  $g_3(\succsim'_1, \succsim'_2, \succsim'_4, \succsim_-^*) \supset N - \{3\}$  for all  $\succsim'_1, \succsim'_2$  and  $\succsim'_4$ . Now let  $\succsim_1^{**}$  be the same as  $\succsim_1^*$ , except that 3 is top-ranked. Likewise, let  $\succsim_2^{**}$  be the same as  $\succsim_2^*$ , except that 3 is top-ranked. In addition, let  $\succsim_4^{**} \in P^\uparrow[3]$ . From above, we have  $g_3(\succsim_1^{**}, \succsim_2^{**}, \succsim_4^{**}, \succsim_-^*) \supset N - \{3\}$ . Let's now consider the  $(3, k)$  marginal mechanism for  $k > 4$ . Of course,  $(3, k) \in I^{3,k}(\succsim_1^{**}, \succsim_2^{**}, \succsim_4^{**}, \succsim_-^*)$  so again we have four possible cases, however there are only two cases which fit the fact that match  $k$  if she top-ranks her, or match, say, 1 if she top-ranks 1 instead. Either 3 is the only dictator in the marginal mechanism, or the mechanism is like panel (A) of figure 10. In the latter case, we have  $f_3(\succsim_1^{**}, \succsim_2^{**}, \succsim'_3, \succsim_4^{**}, \succsim_-^*)$  whenever  $\succsim_k$  top-ranks 3. However, then neither 1 or 2 are getting their top choice. By Maskin monotonicity,

$$f_3(\succsim_1^{**}, \succsim_2^{**}, \succsim'_3, \succsim_4^{**}, \succsim_-^*) = f_3(\succsim_1^*, \succsim_2^*, \succsim'_3, \succsim_4^{**}, \succsim_-^*)$$

<sup>25</sup>Of course, a roommates mechanisms for the agents in  $A$  has a different domain than  $f(\cdot, \succsim_{N \setminus A}^*)$ , but by nonbossiness the ranking of agents in  $N \setminus A$  is immaterial to the mechanism when we restrict attention to  $[P^\uparrow(A)]^A$ .

. Yet if  $\succsim'_3$  top-ranks 4, we get a contradiction. Hence 3 is the dictator in the in the  $(3, k)$  marginal mechanism. So if we switch  $\succsim_k^*$  to  $\succsim_k^{**}$  which is the same, except that 3 is top-ranked, 3's option set is unchanged. Repeating the same analysis shows that

$$g_3(\succsim_{-3}^{**}) \supset N - \{3\}$$

where  $\succsim_l^{**}$  top ranks 3 for all  $l$ . However, repeating our analysis from earlier, we find that in every  $(3, l)$  marginal mechanism 3 is the local dictator, and finally that  $g_3(\succsim_-) \supset N - \{3\}$  for all  $\succsim_-$ . This gives the desired result and proves the theorem for this case.

For the second case in which 1 or 2 is the dictator among the agents 1, 2, 3, 4 when the other agents announce  $\succsim_-^*$ , the strategy will be to reduce this to case 1 by showing that this agent is the dictator among  $N \setminus \{3, 4\}$  for some announcement in which 3 and 4 top-rank each other. Up to relabeling, this is the same as case 1. Since 1 and 2 are symmetric, without loss, suppose the dictator is 1. Then we have

$$f_1(\succsim'_1, \succsim'_2, \succsim'_3, \succsim'_4, \succsim_-^*) = \max_{\succsim'_1} N$$

whenever  $\max_{\succsim'_1} N \in \{1, 2, 3, 4\}$  for all  $\succsim'_1, \succsim'_2, \succsim'_3, \succsim'_4$ . Or, equivalently,  $g_1(\succsim'_2, \succsim'_3, \succsim'_4, \succsim_-^*) \supset \{2, 3, 4\}$  for all  $\succsim'_2, \succsim'_3, \succsim'_4$ . For any  $k$ , if  $\succsim_3 \in P^\uparrow[k]$ , then by previous discussion,  $f_3(\succsim_1^*, \succsim_2^*, \succsim_3, \succsim'_4, \succsim_-^*) = k$  for any  $\succsim_4$ . However, since 1 is the local dictator, we also have that  $f_1(\succsim_1^*, \succsim_2^*, \succsim_3, \succsim'_4, \succsim_-^*) = 2$ . Now if we change  $\succsim_2^*$  to  $\succsim_2^{**}$  by putting 3 on top and we change  $\succsim_3$  to  $\succsim_3^{**}$  by putting 1 on top, then since 1 is the dictator, neither of these changes affect the outcome of 1 and therefore 2, so by Maskin monotonicity,  $f_1(\succsim_1^*, \succsim_2^{**}, \succsim_3^{**}, \succsim'_4, \succsim_-^*) = f_1(\succsim_1^*, \succsim_2^*, \succsim_3, \succsim'_4, \succsim_-^*)$ . However, now consider the  $(1, k)$  marginal mechanism. If  $\succsim_1^{**}$  top-ranks  $k$ , but is otherwise unchanged and  $\succsim_k^{**}$  top-ranks 1, but is otherwise unchanged. then if  $(1, k)$  are not matched at the profile  $(\succsim_1^{**}, \succsim_2^{**}, \succsim_3^{**}, \succsim'_4, \succsim_k^{**}, \succsim_-^*)$ , by Maskin monotonicity, the  $f$  is unchanged. However, this outcome would be Pareto dominated by the match in which 1 and  $k$  are matched, 2 and 3 are matched and all other matches are unchanged. Thus we have that  $(1, k) \in I^{1,k}(\succsim_2^{**}, \succsim_3^{**}, \succsim'_4, \succsim_-^*)$ . Again, this leaves us with four options. However, only 1 as the dictator fits with the fact that, if 1 top-ranks 3, they are matched, and  $k$ 's match is changed. In the three other cases this cannot happen. Therefore, 1 is the dictator in this marginal mechanism. As a consequence we have that  $k \in g_1(\succsim_2^{**}, \succsim_3^{**}, \succsim'_4, \succsim_-^*)$ . Since 2, 3, and 4 cannot affect 1's option set, we have  $k \in g_1(\succsim'_2, \succsim'_3, \succsim'_4, \succsim_-^*)$  for all  $\succsim'_2, \succsim'_3, \succsim'_4$ . Since  $k$  was chosen arbitrarily, we actually have  $g_1(\succsim'_2, \succsim'_3, \succsim'_4, \succsim_-^*) = N - \{1\}$  for all  $\succsim'_2, \succsim'_3, \succsim'_4$ . However, going back to Lemma 7, if 3 and 4 top-rank each other, there is a dictator among the other agents. The fact that 1's option set is  $N - \{1\}$  is only compatible with 1 being that dictator. This gets us back to case 1, and repeating the argument there, we see that 1 always gets her top choice. By the induction assumption, we are done.  $\square$

## A.11 Proof of Lemma 4

Nonbossiness is immediate. Then the result follows from the observation that strategy-proofness and nonbossiness are equivalent to group strategy-proofness, recorded in Proposition 1.  $\square$

## A.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  $f_{\succsim_{-ij}}^{ij}$  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  $|I_{\succsim_{-ij}}^{ij}| = 1$  then  $f$  is single-valued<sup>26</sup> 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  $|I^{12}(\succsim_{-12})| = 2$  and without loss assume  $I^{12}(\succsim_{-12}) = \{a, b\}$ . 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  $\phi_1$  and  $\phi_2$  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  $\mathcal{O}$ . If we let  $\succsim_2^* \in \mathcal{P}^\uparrow[c, a, b]$  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  $I^{1,3}(\succsim_2^*, \succsim_4, \dots, \succsim_n)$ . As before we have two possible mechanisms and in either one, if  $\succsim_3^* \in \mathcal{P}^\uparrow[c, a, b]$  we have  $f(\succsim_1, \succsim_2^*, \succsim_3^*, \succsim_4, \dots, \succsim_n) = a$  if  $a \succ_1 b$  and  $f(\succsim_1, \succsim_2^*, \succsim_3^*, \succsim_4, \dots, \succsim_n) = b$  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  $c \notin \text{im}(f)$  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  $f_{\succsim_{-ij}}^{ij}$ . 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  $f(\succsim'_1, \succsim_2, \dots, \succsim_n)$ . 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  $f(\succsim'_1, \succsim'_2, \dots, \succsim_n)$ . Now we have that  $I^{1,3}(\succsim'_2, \succsim_4, \dots, \succsim_n)$  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  $f(\succsim'_1, \succsim'_2, \succsim'_3, \succsim_4, \dots, \succsim_n) = \max_{\succsim_1} I^{1,2}(\succsim_{-1,2})$ . Continuing in this way gives the desired result. The assumption that  $f$  is surjective implies that 1 is a dictator.

□

<sup>26</sup>To see that  $f(\succsim) = f(\succsim')$ , change one preference at a time. No single change can alter  $f$ , so we get the result.

### A.13 Proof of Theorem 5

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

### A.14 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<sup>27</sup>.  $\square$

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			

### A.15 Proof of Proposition 6

The proofs of these two statements are nearly identical. Let  $\alpha$  and  $\alpha'$  induce  $f$  which is group strategy-proof. Let  $\succsim$  be an arbitrary preference profile and set  $\succsim^0 = \succsim$ . Iteratively define the sequence  $\succsim^0, \succsim^1, \dots, \succsim^N$  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  $\succsim^0, \succsim^1, \dots, \succsim^N$  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  $f^\alpha$  is unchanged along the sequence and again the sequence of  $\tau(\succsim^n)$  follows the allocations in the steps of the constraint-traversing mechanism.  $\square$

### A.16 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  $\alpha'$  be identical to  $\alpha$  except that  $\alpha'(x) = \{i\}$ . We claim that for every preference profile the constraint traversing algorithm using  $\alpha$  and  $\alpha'$  yield the same allocation, so that  $\alpha'$  is implementable

<sup>27</sup>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.

and  $f^\alpha = f^{\alpha'}$ . Of course, the constraint traversing algorithm for  $\alpha$  and  $\alpha'$  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  $\alpha'$  at  $\succsim$  eventually lands on  $x$  at which point  $i$  compromises. After that, the algorithm operates identically to the algorithm under  $\alpha$  at  $\succsim'$ , and therefore yields the same outcome.  $\square$

## A.17 Proof of Proposition 8

Let  $f$  be a constraint-traversing mechanism for the local constraint assignment  $\alpha$ . Pick any proper subset  $M$  of  $N$  and a preference profile  $\succsim_{M^c}$  for the other agents. We must show that  $f_{\succsim_{M^c}}^M$  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 unachievable 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  $\alpha$  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  $\alpha$ , until reaching  $x^*$ , only depends on the top choices of the agents in  $M$ . Thus we may define  $\alpha^*$  likewise on the rest of  $\mathcal{O}^M - I^M$  in a well-defined way. It remains to show that  $\alpha^*$  implementable and that the induced algorithm agrees with  $f_{\succsim_{M^c}}^M$ .

To see this, pick an arbitrary  $\succsim_M$ . If  $\tau(\succsim_M) \in I_{\succsim_{M^c}}^M$ , then the constraint traversing algorithm under  $\alpha^*$  gives the suballocation  $\tau(\succsim_M)$ , which agrees with  $f(\succsim_M, \succsim_{M^c})$  by group strategy-proofness. Otherwise,  $\tau(\succsim_M) \notin I_{\succsim_{M^c}}^M$ . 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  $\succsim_M^2$  by having each agent in  $\alpha^*(\tau(\succsim_M))$  move their top choice to the bottom. Now Maskin monotonicity implies that  $f(\succsim_M^2, \succsim_{M^c}) = f(\succsim_M, \succsim_{M^c})$ . If  $\tau(\succsim_M^2) \in I^M(\succsim_{M^c})$ , we stop. Otherwise, we repeat the process. Continuing in this way, we get a sequence of profiles  $\succsim_M^1, \succsim_M^2, \dots, \succsim_M^n$  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  $\alpha^*$  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  $\alpha^*$ .  $\square$

### A.18 Proof of Theorem 6

We show this in four steps. First, we show that if  $\alpha$  satisfies forward consistency, then any  $\alpha'$  such that for all  $x \in \bar{C}$ ,  $\emptyset \subsetneq \alpha'(x) \subset \alpha(x)$  also implements  $f^\alpha$ . Next, we show that this, along with forward consistency imply that the marginal mechanisms holding a single agents' preferences fixed for  $f^\alpha$  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  $\alpha$  never lands on  $x$  will yield the same result under  $\alpha'$ . Let  $\succsim$  be a preference profile such that the constraint-traversing algorithm under  $\alpha$  eventually lands on  $x$ . Then the sequence of allocations achieved in both algorithms is identical until they both land on  $x$ . Let  $z^0, z^1, \dots, z^p$  be the sequence of allocations after  $x$  in the constraint-traversing algorithm under  $\alpha$  and  $w^0 \rightarrow w^1 \rightarrow \dots, w^q$  be the same sequence for  $\alpha'$ . 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

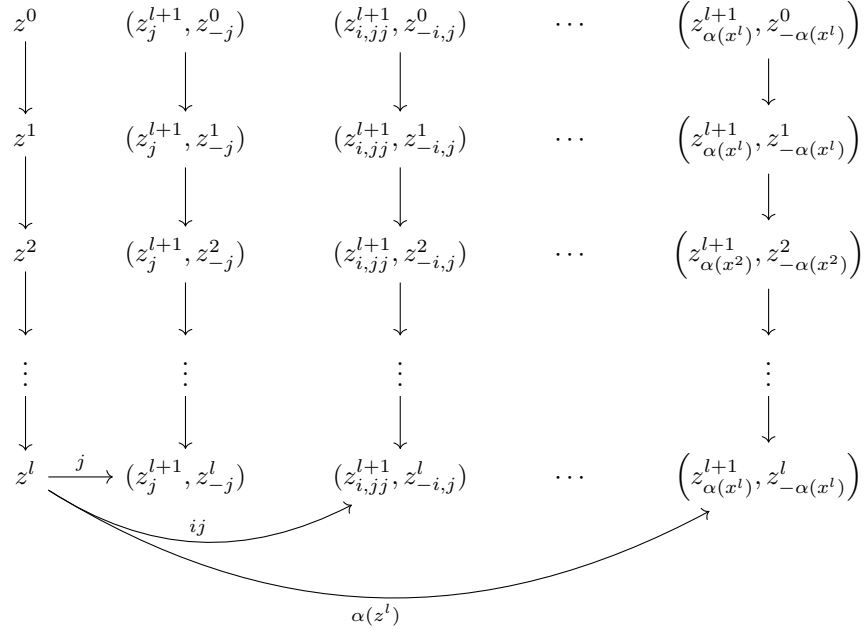
$$w^1 \geq_{PD} z^1 \geq_{PD} w^2 \geq_{PD} z^2 \geq_{PD} w^3 \dots$$

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  $\alpha$  and  $\alpha'$  result in the same outcome at  $\succsim$ . Hence  $f^\alpha = f^{\alpha'}$ . Of course, for any  $\alpha''$  such that  $\emptyset \subsetneq \alpha''(x) \subset \alpha(x)$  on  $\bar{C}$  we can iteratively remove one agent at a time, to get that  $\alpha''$  implements  $\alpha$ .

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,  $\alpha'$  implements  $\alpha$ . 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  $\alpha^*$  implements  $h$ . To do so, we will induct on the number of steps required in the constraint-traversing algorithm under  $\alpha'$ . If  $\succsim$  is a preference profile such that the constraint-traversing algorithm under  $\alpha'$  takes just one step, then  $\tau(\succsim_{-k}, \succsim_k)$  is feasible, so  $h(\succsim) = \tau(\succsim)$  which is the outcome of  $\succsim$  under  $\alpha^*$ . Now assume that the statement holds for all preference profiles which take less than or equal to  $n$  steps under  $\alpha'$ . Let  $\succsim_{-k}$  be a preference profile such that the outcome of  $(\succsim, \succsim_k)$  takes  $n+1$  steps under  $\alpha'$ . Let  $z^0 \rightarrow \dots, z^n$  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  $\alpha'$  is  $z^1 \rightarrow z^n$  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  $\alpha^*$



at  $\succsim_{-k}$ . However, we also have that  $f(\succsim'_{-k}, \succsim_k) = f(\succsim_{-k}, \succsim_k)$  by construction and the outcome of  $\alpha^*$  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  $\alpha^*$ . 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  $(x, z_{-j}^l) \in \bar{C}$  and  $k \in \alpha(x, z_{-j}^l)$ . In particular, we have  $k \in \alpha(z_j^{l+1}, z_{-j}^l)$ . But then we have a monotone path  $(z_j^{l+1}, z_{-j}^l) \xrightarrow{k} (z_j^{l+1}, z_k^{l+1}, z_{-j,k}^l)$ . Furthermore from forward consistency,  $\alpha(x^l) - \{j\} \subset \alpha(z_j^{l+1}, z_k^{l+1}, z_{-j,k}^l)$ . 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  $\alpha^*$ , which is not the first step of the algorithm under  $\alpha'$ . However, the first steps under  $\alpha'$  are just  $k$  compromising. Thus, as shown in the figure, we can use backward consistency to show that 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  $\alpha^*$  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  $\alpha^*$  at  $\succsim_{-k}$ . This gives the desired result.

For step 3, we need to show that  $\alpha^*$  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  $\alpha$  and since  $\alpha(q_k, x) = \alpha'(q_k, x)$ , 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  $\alpha^*$  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  $\alpha^*$  satisfies forward consistency. First we will need a definition. Given a monotone sequence  $z^{(n)} = z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$  we let  $\succsim_{z^{(n)}}$  be the preference profile in which all agents put  $z^0$  as their top choice.  $i_1$  puts  $z^1_{i_1}$  as her next best choice and so on. We will also need the following proposition:

**Proposition 10.** *If  $\eta$  satisfies forward and backward consistency and  $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$  is a monotone sequence for  $\eta$  and  $i_p \neq i_l$  with  $l < p$  then there is a monotone sequence  $w^{(n)}$  starting at  $(z^0_{-i_p}, z^p_{i_p})$ , 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  $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_{n+1}} z^{n+1}$  be a monotone sequence such that  $i_{n+1}$  hasn't compromised before. By backward consistency  $\eta(z^0_{-i_{n+1}}, z^{n+1}_{i_{n+1}})$  intersects  $\{i_1, \dots, i_n\}$ . Suppose specifically that  $i_l \in \eta(z^0_{-i_{n+1}}, z^{n+1}_{i_{n+1}})$ . By the induction hypothesis, there is a monotone sequence  $w^{(n)}$  of length  $l-1$  which starts at  $(z^0_{-i_l}, z^l_{i_l})$  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  $(z^0_{-i_l, i_{n+1}}, z^l_{i_l}, z^{n+1}_{i_{n+1}})$  to  $z^{n+1}$ . However, this can now just be the second step of a new monotone sequence that starts at  $(z^0_{-i_{n+1}}, z^{n+1}_{i_{n+1}})$ . This gives the desired result.  $\square$

Now, we are ready to show that  $\alpha^*$  satisfies forward consistency. Given a monotone sequence  $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$  under  $\alpha^*$ , 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  $\{i_1, \dots, i_n\} \cup \{k\}$ . 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  $\{i_1, \dots, i_n\}$ . 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.  $\square$

## A.19 Proof of Proposition 9

Suppose that  $\alpha$  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  $z^0 \xrightarrow{i_1}$

$z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$  in which an agent  $k$  compromises  $|\mathcal{O}| - 1$  times. For each  $i$ , let  $\succsim_i$  top-rank  $z_i^0$ , put  $z_i^1$  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  $\alpha'$  is the same as under  $\alpha$ . However the constraint-traversing algorithm for  $\alpha'$  under  $\succsim$  runs out of  $k$ 's possible allocations. Hence  $\alpha'$  is not implementable,  $\alpha$  is not implementable, a contradiction.

Now suppose that  $\alpha$  satisfies forward and backward consistency and here is no monotone path  $z^0 \xrightarrow{i_1} z^1 \xrightarrow{i_2} \dots \xrightarrow{i_p} z^p$  in which an agent  $i$  compromises  $|\mathcal{O}| - 1$  times. Suppose by way of contradiction, that  $\alpha$  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  $w^0 \rightarrow w^1 \dots \rightarrow w^q$  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.  $\square$

## References

- 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.
- 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.
- 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.
- 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.
- PYCIA, M., AND U. ÜNVER (2017): “Incentive Compatible Allocation and Exchange of Discrete Resources,” *Theoretical Economics*, 12, 287–329.
- ROTH, A. E. (2002): “The economist as engineer: Game theory, experimentation, and computation as tools for design economics,” *Econometrica*, 70(4), 1341–1378.
- 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.
- 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.

## S Supplemental Appendix

### S.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.

1&2	1	1
2		
2		

(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.

1&2	1	1
2	1 or 2	
2		1 or 2

(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.

1&2	1	1
2		
2		2

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

1&2	1	1
2		
2		2

(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.

1&2	1	1
2		
2		2

(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.

1&2	1	1
2	1 or 2	
2		2

1 or 2	2	
1	1&2	1
	2	2

2		2
	2	2
3	3	2&3

(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.

1&2	1	1
2	1	
2		2

	3	
	1&3	
	3	2

	1	
3	3	2&3

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

1&2	1	1
2	1	
2		2

1	3	
1	1&3	
	3	2

2		
	1	3
3	3	2&3

(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.

1&2	1	1
2	1	
2		2

1	3	
1	1&3	3
	3	2

2		2
	1	3
3	3	2&3

(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.

1		



(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,

1		
2		
3		



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

1	1	1
2		
3		

1		

1		

(13) The highlighted squares follow from backward consistency.

1	1	1
2	2	
3		

1		

1		
3		

(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.

1	1	1	1			1		
2	2		1					
3		1				3		

(15) The highlighted square in the second panel cant 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.

1	1	1	1			1		
2	2		1	2 or 1&2				
3		1				3		3 or 1&3

(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) cant have a 3 because backward consistency would require a 3 in the squares (b,x,b) which cant be efficient, since 1 is already listed at (b,a,b).

1	1	1	1	2		1		3
2	2		1	2				3
3		1		2		3		3

(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.

1	1	1	1	2		1		3
2	2		1	2	3			3
3		1		2	2	3		3

(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.

1	1	1	1	2		1		3
2	2		1	2	3		2	3
3		1		2	2	3	3	3

(19) The square (c,b,c) has to be labeled "3" by backward consistency. Then the other square cant 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

1	1	1	1	2		1		3
2	2		1	2				3
3		1		2	1	3		1&3

(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.

1	1	1	1	2		1		3
2	2		1	2				3
3		1		2	1	3	1	1&3

(21) The highlighted square cant 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).

1	1	1	1	2		1		3
2	2		1	2			1	3
3		1		2	1	3	1	1&3

(22) The highlighted square cant have a 1 or a 2 because both lead to immediate violations of backward consistency.

1	1	1
2	2	
3		1

1	2	
1	2	3
	2	1

1		3
	1	3
3	1	1&3

1	1	1
2	2	
3		1

1	2	
1	1&2	1
	2	1

1		3
		3
3		1&3

(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 follows from forward consistency.

1	1	1
2	2	
3		1

1	2	
1	1&2	1
	2	1

1		3
	1	3
3	1	1&3

(25) The final squares have to be filled out as follows. If either has a "3" we get a 3 step monotone path which cant satisfy backward consistency. The same happens if either is labeled "2." This labeling follows satisfies both consistency conditions. Hence we get a GSP, PE mechanism.

## S.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).

		1			2			3

(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
								3
		1			2	3	3	3

(3) The highlighted squares are implied by backward consistency

								3
								3
		1		2	2	3	3	3

(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).

		1						3
								3
		1		2	2	3	3	3

(5) The highlighted square cant be labeled 2 or 3 because both give immediate violatins of backward consistency.

		1				3		3
								3
		1		2	2	3	3	3

(6) The highlighted square cant have a 1 or a 2 since backward consistency would imply that (a,a,a) should have the same label in both cases.



		1	3			3		3
								3
		1		2	2	3	3	3

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

		1	3			3		3
			3					3
		1		2	2	3	3	3

(8) The highlighted square cant have a 2 since eventually another agent will have to compromise leading to a violation of backward consistency. It cant have a 1 because this would lead to an immediate violation of backward consistency.

		1	3			3		3
			3		3			3
		1		2	2	3	3	3

(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.

		1	3			3		3
			3	2&3	3			3
		1		2	2	3	3	3

(10) This square cant 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.

		1	3			3		3
	2		3	2&3	3		2	3
		1		2	2	3	3	3

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

		1	3	3		3		3
	2		3	2&3	3		2	3
		1		2	2	3	3	3

(12) The highlighted square cant have a 1 or a 2 since both lead to immediate violations of backward consistency.

	3	1	3	3		3		3
	2		3	2&3	3		2	3
		1		2	2	3	3	3

(13) The highlighted square cant have a 1 or a 2 since both lead to immediate violations of backward consistency.

	3	1	3	3		3		3
	2		3	2&3	3		2	3
3		1		2	2	3	3	3

(14) The highlighted square cant have a 1 or a 2 since both lead to immediate violations of backward consistency.

	3	1		3	3		3		3
2 or 3	2			3	2&3	3		2	3
3		1			2	2	3	3	3

(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.

					1&3				
									1&2

(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.

				3				2
	1			1&3			1	2
				3		1	1	1&2

(17) The implications of forward consistency.

				3		1 or 2		2
	1			1&3			1	2
				3		1	1	1&2

(18) The highlighted square cant have a 3 since it gives two different violations of backward consistency.

			1	3		1 or 2		2
	1			1&3			1	2
				3		1	1	1&2

(19) The highlighted square cant have a 3 because either way we fill out (a,a,c) we get a violation of backward consistency. It cant have a 2 because of the immediate violation of backward consistency.

			1	3		1 or 2		2
	1		1	1&3			1	2
				3		1	1	1&2

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

			1	3		1 or 2		2
1	1		1	1&3			1	2
				3		1	1	1&2

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

			1	3		1 or 2		2
1	1		1	1&3			1	2
1				3		1	1	1&2

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

			1	3		1 or 2		2
1	1		1	1&3			1	2
1		1		3		1	1	1&2

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

			1	3		1 or 2		2
1	1		1	1&3			1	2
1		1		3	1	1	1	1&2

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

			1	3		1 or 2		2
1	1		1	1&3	1		1	2
1		1		3	1	1	1	1&2

(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.

		1	1	3		1 or 2		2
1	1		1	1&3	1		1	2
1		1		3	1	1	1	1&2

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

	1 or 3	1	1	3		1 or 2		2
1	1		1	1&3	1		1	2
1		1		3	1	1	1	1&2

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

				2				2
			1	1&2	1			2
				2		1	1	1&2

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

				2				2
			1	1&2	1			2
				2		1	1	1&2

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

	1	1	1 or 2	2		1 or 2		2
1	1		1	1&2	1		1 or 2	2
1		1		2	1 or 2	1	1	1&2

(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