Strategy-Proof Exchange under Trichotomous Preferences

Vikram Manjunath[*] Alexander Westkamp[†]

September 17, 2020

Abstract

We study the balanced exchange of indivisible objects without monetary transfers when agents may be endowed with (and consume) more than one object. We propose a natural domain of preferences that we call *trichotomous*. In this domain, each agent's preference over bundles of objects is responsive to an ordering over objects that has the following three indifference classes, in decreasing order of preferences: desirable objects, objects that she is endowed with but does not consider desirable, and objects that she neither is endowed with nor finds desirable.

For this domain, we define a class of individually rational, Pareto-efficient, and strategy-proof mechanisms that are also computationally efficient.

JEL classification: C78, D82

Keywords: individual rationality, Pareto-efficiency, strategy-proofness, indivisible goods, multi-unit demand

1 Introduction

We study the problem of reallocating indivisible objects without monetary transfers.[1] Unlike much of the earlier work on this problem, we consider situations where each agent may be

[*]Department of Economics, University of Ottawa; e-mail: vikram@dosamobile.com.

[†]Department of Management, Economics and Social Sciences, University of Cologne; e-mail: westkamp@wiso.uni-koeln.de.

[1]See the literature following Shapley and Scarf [1974].

endowed with and consume more than one object. Our contribution is to define a natural domain of preferences and a class of individually rational, Pareto-efficient, strategy-proof mechanism, and computationally efficient. Our positive result is in marked contrast with the impossibility results that abound in the literature on multi-object exchange without monetary transfers [Sönmez, 1999, Biró et al., 2018].

Our interest in these exchange problems stems from our search for a solution to the problem of shift-reallocation. Millions of people in many different professions, from physicians to retail workers, engage in shift work. Shift plans are often made months in advance and scenarios like the following are common: A medical practice consists of four specialist consultant doctors A, B, C, and D. This practice is responsible for ensuring that emergency medical services in their specialty are available to a given hospital at all times. That is, each week, one of the four doctors is to be designated as being *on call*. Being on call is not desirable for these doctors. However, it is necessary for their practice to maintain privileges at the hospital. Since it is a chore that they must perform, in the interest of fairness, they agree to share the weeks equally so that each member of the group is on call every fourth week.[2] To facilitate planning, the call schedule is made six months at a time, taking the doctors' preferences into consideration. For instance, the schedule from January 1 to June 30 is announced in December, and is created on the basis of the doctors' preferences as of December. Thus, on January 1, each of the four doctors is responsible for their assigned weeks until June 30. While this initial assignment may be Pareto-efficient with regards to the doctors' December preferences, six months is a long time. At some later point, say the beginning of March, there may be scope for re-optimization based on current preferences. It may well happen that Dr. A would like to attend a conference the week of April 5, Dr. D would like to help at a clinic in a remote area on June 16, and Dr. C would like to go on vacation on May 23. If each of these doctors is obliged to be on call for the respective week, a three-way trade could improve welfare in regards to current (as of March) preferences. We are interested in the design of mechanisms to identify such trades optimally and provide agents with incentives to reveal their private information about scheduling conflicts truthfully.[3,4]

As the above example illustrates, we are interested in the problem of *re-allocating* shifts,

[2]In the case of medical residents, such a restriction would not only be for reasons of fairness, but also for training purposes.

[3]Note that the re-optimization at the beginning of March in the above example is a static problem. We do not consider the issue of optimally timed matching as in Akbarpour et al. [2020].

[4]To the best of our knowledge, in practice these types of reassignments are typically arranged manually. As a result, more complex trades (three-way or larger) are often unrealized, leaving gains from trade on the table. Software solutions for managing call schedules facilitate these types of trades manually as well. The initial schedule (endowment) is similarly manually constructed. See the following for more on shift exchange in Microsoft's Teams, which is used at many hospitals: https://support.microsoft.com/en-us/office/manage-your-shifts-in-shifts-50920ad2-a0e7-4f4a-aeb8-adf57d3b0c88.

or indivisible objects more generally, from fixed endowments rather than designing the initial schedule itself.[5] This problem is relevant for many workers: given a fixed schedule, a worker may wish to engage in other activities (like work for other firms, further training or education, vacation) that are incompatible with her currently assigned responsibilities; at the same time, workers will have typically already made some commitments that limit the set of new shifts that they can take on. As long as there are possibilities of shocks to preferences or opportunities to make desirable commitments over the duration of the schedule, workers are bound to find themselves in situations where there are gains from trading pre-assigned shifts.

We define the domain of *trichotomous* preferences, which suits applications like shift exchange, as follows. Each agent's preference over bundles of objects is responsive[6] to an ordering over objects that that has three indifference classes:

1. Desirable objects. This set may include some of her endowed objects. In the shift exchange application, these correspond to shifts that she has no scheduling conflict with.

2. Objects that she is endowed with but does not find desirable. These correspond to shifts that she is already assigned and is therefore obliged to fulfill, but would like to trade away so that she may make other plans.

3. Objects that she is neither endowed with nor finds desirable. These are shifts that she cannot take on due to other prior commitments.

We focus on situations in which endowments are commonly known and exchange is balanced in the sense that each agent ends up with the same number of goods as she is endowed with. In the context of shift exchange, this mean that all participants know the initial schedule of shifts and that a worker's total workload is fixed at the number of shifts in the initial schedule.[7]

Trichotomous preferences are a natural restriction since we consider balanced exchange. If an agent is endowed with k objects and there are a total of N objects, then there are

[5] Without the requirement of respecting some lower bounds on welfare, the problem of designing an initial schedule is decidedly simpler: one could use a simple version of serial dictatorship where, one after another, agents select their most preferred bundles from what is left after those selecting before.

[6] A preference ordering over bundles is responsive to an ordering of individual objects, if, for each pair $o, p \in O$ and each $B \subseteq O \setminus \{o, p\}$, the former ranks $B \cup \{o\}$ strictly higher than $B \cup \{p\}$ if and only the latter ranks o strictly higher than p. So if an agent's preference over bundles is responsive, how she compares a one-for-one trade does not depend on which other trades the agent participates in.

[7] The restriction to balanced exchange is particularly meaningful for the shift exchange application as it may be imposed due to training purposes (as in the case of medical residents), fairness considerations (sharing equally the burden of taking on call), or terms of employment contracts (particularly for unionized employees who cannot work more or fewer shifts than prescribed by a collective bargaining agreement). See Dur and Ünver [2019], Andersson et al. [forthcoming] and Biró et al. [2018] for more on balanced exchange.

$\binom{N}{k}$ different bundles she could end up with. Even for very modest sized problems, this is an unreasonably large number: with just ten agents, each of whom is endowed with only three objects, there are 4060 different bundles that each of them could consume. That it is impractical to expect an agent to even know, much less communicate, preferences over such a large set has been previously noted in the literature [Milgrom, 2009, 2010, 2011, Othman et al., 2010, Budish and Kessler, 2018]. Asking each agent to submit a rank-order list over objects is far more tractable. This is the approach taken, for instance, in the National Resident Matching Program: hospitals have preferences over bundles of doctors but submit a ranking of individual doctors. If hospitals' preferences are responsive, this simplification in the message space does not change the set of equilibrium outcomes compared to when hospitals may actually submit full rankings over bundles [Milgrom, 2010, Perez-Richet, 2011]. Even if we restrict preferences to be responsive, the total number of objects N may be large. Ranking all of the objects, let alone communicating them to the mechanism, may be difficult. Indeed, there is evidence in the marketing literature that the larger the set a consumer chooses from, the less intense her preferences may be [Greenleaf and Lehmann, 1995, Dhar, 1997, Iyengar and Lepper, 2000, Chernev, 2003]. So we have posited that each agent partitions these objects into three sets as we have described above.

The objects of our study are mechanisms that associate every profile of preferences with an allocation. We are interested in mechanisms that satisfy the following desiderata:

1. *Individual rationality*, which says that every agent ends up at least as well of as keeping her endowment.

2. *Pareto-efficiency*, which says that no further reallocation can make any agent better off without harming another.

3. *Strategy-proofness*, which says that no agent can ever profit from lying about her preferences, regardless of what the other agents report. In other words, it is a weakly dominant strategy for each agent to truthfully report her preferences.

While we define these three properties for general preference domains, we define a stronger form of individual rationality for trichotomous preferences. *Component-wise individual rationality* says that for each agent, each component of the bundle that she receives is at least as desirable as some object in her endowment. In other words, her bundle ought to be composed only of objects in the top two indifference classes. On the domain of trichotomous preferences, component-wise individually rational mechanisms have the advantage of being "simple" in the sense that the only information an agent needs to convey is the set of objects that she finds desirable. On the other hand, if a mechanism is merely individually rational, it may depend on more information about how an agent compares bundles. After all, the desirable set alone says nothing about how she compares, for instance, a bundle comprised of objects in both the first and third indifference class against her endowment. We show that ignoring such information does not come at the expense of Pareto-efficiency (Proposition 1).

Our main contribution to define a new class of mechanisms satisfying these desiderata.[8] A *Component-wise Individually Rational Priority* (CIRP) mechanism is parameterized by a fixed ordering of the agents, independent of their reports, and picks a final allocation of goods in accordance with the following sequential process:

- In the first step, find all of the allocations that maximize the first agent's preferences from among all component-wise individually rational allocations. This results in a "promised" welfare level for the first agent.

- In the t^{th} step, maximize the t^{th} agent's preferences, subject to the promises made to the first $t-1$ agents, from among all component-wise individually rational allocations. This results in a "promised" welfare level for the t^{th} agent.[9]

- After the last step, reallocate the objects in a way that the promise made to each agent is respected. First, there is necessarily at least one way to do this. Second, as we show, if there are two different re-allocations that achieve this, then every agent is indifferent between them.

A CIRP mechanism selects an outcome that meets the promises made to all of the agents in this process.[10] We show that, given any ordering of the agents, a CIRP mechanism is component-wise individually rational, Pareto-efficient, and strategy-proof. Component-wise individual rationality is by definition. However, unlike the familiar versions serial dictatorship that are common in the literature, Pareto-efficiency and strategy-proofness are not trivial consequences of sequentially maximizing agents' welfare. Extra care is needed to show that the additional constraints of component-wise individual rationality do not prevent the CIRP process from reaching the Pareto-frontier. Establishing strategy-proofness is not obvious since the report of an agent alters the set of component-wise individually rational allocations and

[8]For mechanisms that solve the unit endowment problem under general weak preferences, see Jaramillo and Manjunath [2012], Alcalde-Unzu and Molis [2011], Aziz and De Keijzer [2012], and Saban and Sethuraman [2013].

[9]Note that this does not pin down exactly which objects the t^{th} agent, or any agent before her, receives. It merely fixes the *number* of desirable objects that she will receive in the end.

[10]If each agent were endowed with a single object, this algorithm would essentially amount to the priority algorithm that has been studied in the kidney exchange literature [Roth et al., 2005, Hatfield, 2005, Sönmez and Ünver, 2014].

can thereby affect the outcomes for all agents and not only those who come after her.[11] It turns out that—because exchanges are balanced and we consider trichotomous preferences—one can arrive at any allocation via some set of one-for-one trading cycles from the endowment. This means that if a component-wise individually allocation is not Pareto-efficient, there has to be a one-for-one trading cycle that improves it in the Pareto sense. CIRP ensures that there is no such cycle. To show that CIRP is strategy-proof, we show that if an agent adds an object to or removes an object from her desirable set, she can add or remove, respectively, at most one trading cycle. Given that the trichotomous domain rules out complementarities between the objects, no other cycles can be affected.

Apart from their desirable allocative and incentive properties, we also show that CIRP mechanisms are computationally efficient. More precisely, taking agents' desirable sets and a priority order over the agents as inputs, a CIRP allocation can be computed in $O(mn^3 \log(k))$ time, where m is the total number of objects, n is the number of agents, and k is the largest number of objects an agent may be endowed with.

While the characterization of a maximal domain of preferences where our desiderata are compatible is beyond the scope of this paper, we we discuss to which extent our approach does or does not extent beyond the trichotomous domain.

Related Literature

This work contributes to two strands of literature. The first strand is that which extends the model of reallocating indivisible objects [Shapley and Scarf, 1974] to the case where agents may consume more than one. The second is the literature on allocating objects under very coarse (or "dichotomous" [Bogomolnaia and Moulin, 2004]) preferences.

The model of reallocating objects among agents, each of whom is endowed with and consumes a single object, has many nice properties. Notably, the core is single valued [Roth and Postlewaite, 1977] and it defines a strategy-proof mechanism [Roth, 1982]. However, with only mild restrictions on preferences, these results do not carry over when agents may be endowed with or may consume more than one object [Sönmez, 1999, Konishi et al., 2001, Biró et al., 2018]. Even other strategic properties such as immunity to manipulation via altering of endowments is generally incompatible with individual rationality and Pareto-efficiency [Klaus et al., 2006, Atlamaz and Klaus, 2007]. If we weaken the strategic requirement to say that successful manipulations are computationally intractable to compute [Pini et al., 2011], then there are adaptations of the top-trading-cycles algorithm that satisfy

[11] In the familiar version of serial dictatorship described in Footnote 5, without being subjected to individual rationality constraints, an agent's report can only influence the outcomes for agents who pick after her. It is precisely this feature that makes strategy-proofness straightforward.

this property while maintaining individual rationality and Pareto-efficiency [Fujita et al., 2015, Sikdar et al., 2017, 2018, Phan and Purcell, 2018]. The compatibility of individual rationality, Pareto-efficiency, and strategy-proofness in our model is driven by balancedness and the restricted domain of preferences.

The literature on the "multi-type" version of this model [Moulin, 1995] considers objects that are partitioned into various types and exchange only occurs within a given type. When the objects are not disposable, this implies balanced exchange. However, because each object can only be traded within its own type-market, the only interaction between objects of different types is through the agents' preferences. This is a major difference from our approach, where we place no such restrictions on trades. To the best of our knowledge, while preferences over bundles have been restricted to be separable [Klaus, 2008, Anno and Kurino, 2016] or lexicographic [Monte and Tumennasan, 2015], restrictions on preferences over objects of the same type—that is, marginal preferences—remain unstudied as a way around the incompatibility of our three main desiderata.

Among recent works that do not restrict trade across types of objects, Biró et al. [2018] consider a model where each agent is endowed with some fixed capacity for supplying indistinguishable copies of some object. Like the current work, attention is restricted to balanced exchange. However, preferences are responsive to strict, as opposed to trichotomous, orderings over the objects. Biró et al. [2018] characterize the capacity vectors for which individual rationality, Pareto-efficiency, and strategy-proofness are compatible. For capacity vectors in this class, it is shown that a variant of the top-trading cycles mechanism is the unique mechanism that satisfies the three desiderata. For capacity vectors outside of this class, Biró et al. [2018] study the incentive and efficiency properties of different variations of the top-trading-cycles algorithm.

For the two-sided one-to-one matching problem, the analog of our domain—where each agent divides the set of partners into desirable and undesirable ones and is indifferent among those in the same group—results in these properties being compatible as well [Bogomolnaia and Moulin, 2004]. This domain of preferences is a natural fit for modelling kidney exchange problems: compatible donors are desirable and incompatible ones are not. Whether the restriction on exchanges is that they be pairwise—with compatible pairs [Sönmez and Ünver, 2014] or without [Roth et al., 2005]—or more general [Hatfield, 2005], a priority mechanism much like IRP turns out to be individually rational, Pareto-efficient, and strategy-proof. While one might think that these results are driven by the fact that there are only two welfare states for each agent in these kidney exchange models—matched or unmatched—our results suggest that this is not so: the IRP approach is fruitful even when agents preferences are monotonic in the number of desirable objects that they receive.

The only other such extension of Bogomolnaia and Moulin [2004], as far as we know, to a model of balanced exchange in which agents are endowed with and consume multiple objects is by Andersson et al. [forthcoming]. Their independent and concurrent work differs from ours in some important ways. On one hand, the domain that these authors work with is significantly more restricted than our trichotomous domain: like in Biró et al. [2018] agents are endowed with identical copies and each agent finds any bundle containing an object worse than her endowment to be unacceptable. On the other hand, Andersson et al. [forthcoming] demonstrate the existence of individually rational and strategy-proof mechanisms that satisfy a more demanding efficiency requirement—maximization of the number of objects that are traded—than Pareto-efficiency. Since there is a trade-off between generality of the preference domain and how demanding an objective can be satisfied, our results are logically independent from theirs.

Organization of the paper

In Section 2 we introduce the model and define the trichotomous domain. In Section 3 we define our mechanisms. In Section 4 we show that our mechanisms are strategy-proof. In Section 5 we show how to compute the output of our mechanisms in polynomial time. In Section 6 we discuss the extension of our approach to non-trichotomous preference domains. All proofs are in the Appendix.

2 The Model

Let $I = \{1, \ldots, n\}$ be a set of n agents and O be a finite set of objects. Each agent $i \in I$ is endowed with a non-empty set of objects $\Omega_i \subseteq O$. We assume throughout that $\Omega_i \cap \Omega_j = \emptyset$ whenever $i \neq j$ and that $O = \cup_{i \in I} \Omega_i$.[12] For the shift exchange application, the endowment represents the pre-arranged schedule. We assume that these endowments are known and fixed at the profile $\Omega = (\Omega_i)_{i \in I}$. In the context of shift exchange, the assumption of known endowments is reasonable since one would expect firms and workers to know who is supposed to work when, once an initial assignment of shifts has been determined (and no shift exchanges were to take place).

A *matching* is a mapping $\mu : I \to 2^O$ that satisfies the following two requirements:

1. For any pair of distinct agents $i, j \in I$, $\mu(i) \cap \mu(j) = \emptyset$.

[12] In our model, each agents has to end up with exactly the same number of objects as she was endowed with. Given this restriction, it is easy to accommodate the case where some objects are not in the endowment of any agent ($\cup_{i \in I} \Omega_i \subsetneq O$) by introducing extra (dummy) agents, who are initially endowed with the objects in $O \setminus \cup_{i \in I} \Omega_i$. Details are available upon request.

2. For any agent $i \in I$, $|\mu(i)| = |\Omega_i|$.

The endowment, Ω, itself corresponds to a matching. With slight abuse of notation, we denote the matching that assigns to each agent her endowment by Ω. Let $\mathcal{M}$ be the set of all matchings. Our definition of a matching requires the exchange of objects to be *balanced* in the sense that each agent ends up with the same number of objects as she was endowed with. For applications like shift exchange, this type of restriction is natural as contracts often specify a fixed number of shifts per worker.

Due to balancedness, the set of bundles that an agent $i \in I$ may consume at some matching in $\mathcal{M}$, her *consumption set*, is

$$X_i = \{B_i \subseteq O : |B_i| = |\Omega_i|\}.$$

Each agent $i \in I$ has a weak preference relation $\succsim_i$ over X_i.[13] Let $\mathcal{R}_i$ be the set of all such preference relations. Let $\mathcal{R} = \times_{i \in I} \mathcal{R}_i$ be the set of all preference profiles.

Given a preference profile $\succsim = (\succsim_i)_{i \in I} \in \mathcal{R}$ and a matching $\mu \in \mathcal{M}$, μ is *individually rational under* $\succsim$ if each agent finds her assignment to be at least as good as her endowment—that is, for each $i \in I$, $\mu(i) \succsim_i \Omega_i$. Note that there is always at least one individually rational matching: the endowment itself.

Next, $\mu' \in \mathcal{M}$ *Pareto-improves* upon $\mu \in \mathcal{M}$ under $\succsim$ if each agent likes μ' at least as well as μ and some agent likes μ' strictly better than μ—that is, for each $i \in I$, $\mu'(i) \succsim_i \mu(i)$ and, for some $i \in I$, $\mu'(i) \succ_i \mu(i)$. If there is no μ' that Pareto-improves upon μ, then μ is *Pareto-efficient.*

A *(Cartesian-)domain* of preferences is a subset of the permissible preference profiles that can be expressed as a cartesian product—that is, a domain is $\mathcal{D} \subseteq \mathcal{R}$ such that $\mathcal{D} = \times_{i \in I} \mathcal{D}_i$ and, for each $i \in I$, $\mathcal{D}_i \subseteq \mathcal{R}_i$. Given a domain $\mathcal{D}$, a *mechanism on* $\mathcal{D}$ is a mapping $\varphi : \mathcal{D} \to \mathcal{M}$ that associates a matching to each profile of preferences in $\mathcal{D}$. Given a profile $\succsim \in \mathcal{D}$, we denote the set of objects that i receives under mechanism φ by $\varphi_i(\succsim)$.

A mechanism is *individually rational* if it always (for any profile of preferences in its domain) selects an individually rational matching, and *Pareto-efficient* if it always selects a Pareto-efficient matching. Finally, a mechanism is *strategy-proof* if no agent can ever benefit by reporting a preference that is different from her true one—that is, mechanism φ on $\mathcal{D}$ is strategy-proof if there do not exist $\succsim \in \mathcal{D}$, $i \in I$, and $\hat{\succsim}_i \in \mathcal{D}_i$ such that $\varphi_i(\hat{\succsim}_i, \succsim_{-i}) \succ_i \varphi_i(\succsim)$.

As discussed in the introduction, it is well known that the three key desiderata of individual rationality, Pareto-efficiency, and strategy-proofness are incompatible on $\mathcal{R}$. We now proceed to define a simple domain of preferences that only asks agents to classify others' endowed

[13]In other words, $\succsim_i$ is a complete, transitive, and reflexive relation over X_i.

objects as being desirable or not. In the context of shift exchange, a desirable object can be interpreted as a shift for which the agent does not have a scheduling constraint. As we will see later, we obtain positive results on this domain. In Section 6, we discuss to what extent positive results can be obtained outside our preference domain.

Our preference domain places two restrictions on agents' preferences over bundles. Fix an agent $i \in I$. First, we assume that i has *responsive* preferences over bundles of objects—that is, for agent i there exist a ranking $\succeq_i \in \mathcal{R}_i$ and weak preference relation $\succsim_i^*$ over O such that, for any pair of objects $o, p \in O$ and any set $B \subseteq O \setminus \{o, p\}$, $B \cup \{o\} \succsim_i B \cup \{p\}$ if and only if $o \succsim_i^* p$. Second, we assume that agent i's preference relation over individual objects has a particularly simple form that we call *trichotomous*. Formally, a weak preference relation $\succsim_i^*$ over O is trichotomous for i if there is a set of *desirable* objects $A_i \subseteq O$ such that $\succsim_i^*$ ranks every desirable object in its top indifference class, every endowed object that is not desirable in its second indifference class, and all remaining objects in its third indifference class. That is, $\succsim_i^*$ satisfies the following:

1. $o \sim_i^* p$ for all pairs o, p such that either $\{o, p\} \subseteq A_i$, $\{o, p\} \subseteq \Omega_i \setminus A_i$, or $\{o, p\} \subseteq O \setminus (A_i \cup \Omega_i)$
2. $o \succ_i^* p$ for all pairs o, p such that either $o \in A_i$ and $p \in O \setminus A_i$, or $o \in \Omega_i$ and $p \in O \setminus (A_i \cup \Omega_i)$

In the context of shift exchange, objects in A_i represent shifts for which the agent i does not have any scheduling conflicts, objects in $\Omega_i \setminus A_i$ represent shifts that are currently assigned to i and are in conflict with some other activities (like additional paid labor for other firms, further training/education, vacation) that i would like to commit to, and of objects in $O \setminus (A_i \cup \Omega_i)$ are shifts that i is not able to take on due to existing scheduling conflicts.[14] Our motivation for the distinction between shifts in $\Omega_i \setminus A_i$ and those in $O \setminus (A_i \cup \Omega_i)$ is that shifts in the latter set conflict with other binding commitments of doctor i, while shifts in the former conflict with tentative commitments. Put differently, agent i is *obliged* to consume the objects in Ω_i if she is not able to trade them for something else. Objects in $\Omega_i \cap A_i$ are those shifts that that i is assigned but has no scheduling conflicts.[15]

The following definition summarizes our assumptions about agents' preferences.

Definition 1. For each agent $i \in I$, $\succsim_i \in \mathcal{R}_i$ is *trichotomous* if it is responsive to a trichotomous ordering over O. Let $\mathcal{T}_i$ be the set of all trichotomous preferences for i and $\mathcal{T}$ the set of all trichotomous preference profiles.

[14] Our results hold even if we allow agent $i \in I$ to rank objects in $I \setminus (A_i \cup \Omega_i)$ in any way, as long as they are ranked below objects in $A_i \cup \Omega_i$.

[15] Objects in $\Omega_i \cap A_i$ correspond to the "compatible pairs" in the kidney exchange literature.

Before proceeding, we discuss the relationship of our trichotomous preference domain to other approaches in the related literature. For that purpose, note first that we allow for the case where $\Omega_i \cap A_i \notin \{\emptyset, \Omega_i\}$, that is, the case where some of i's endowed objects are desirable, while others are not. Hence, we need information about endowments *and* desirable sets of objects in order to describe agents' preferences.[16] This is a distinct feature of our model compared with the *dichotomous* domain of Bogomolnaia and Moulin [2004] and a key reason for calling our domain trichotomous.[17] We also allow for the possibility that a non-empty strict subset of j's endowment Ω_j is desirable for $i \neq j$—that is, $A_i \cap \Omega_j \neq \emptyset$ and $\Omega_j \nsubseteq A_i$. This is the first important way that our's is different from the independent contribution of Andersson et al. [forthcoming]. There, agents are assumed to find desirable either none or all of the objects another agent brings to the market. This assumption is hard to justify for applications like shift exchange as it is clearly possible that i is able to take on some shifts initially assigned to j but not others. The second critical distinction between our domain and that of Andersson et al. [forthcoming] is that we do not require i to find any bundle that gives her one or more objects in $O \setminus (A_i \cup \Omega_i)$ to necessarily be worse than her endowment Ω_i. In our domain, agents may be willing to trade off receiving desirable objects against receiving undesirable ones. For instance, an agent may be willing to accept an object that she is neither endowed with nor finds desirable as long as she receives at least a certain number of desirable objects. The domain of Andersson et al. [forthcoming] rules out such preferences.

While agents with trichotomous preferences may sometimes be willing to participate in exchanges that leave them with new objects that are not desirable, it seems natural to try to prevent such exchanges as much as possible. For example, in the context of shift exchange, we would ideally like to have a mechanism that, if at all possible, does not create any new scheduling conflicts (by assigning some agents to new shifts that they do not find desirable). We now define a stronger form of individual rationality for trichotomous preferences that requires that each agent only be assigned objects that she either finds desirable or that she was endowed with. Importantly, this strengthening is not in conflict with Pareto-efficiency, as we show below. Formally, fix a profile of trichotomous preferences $\succsim \in \mathcal{T}$ and let $A = (A_1, \ldots, A_n)$ be the associated profile of desirable sets. We say that a matching μ is *component-wise individually rational*, if, for each $i \in I$, $\mu(i) \subseteq A_i \cup \Omega_i$. Component-wise individual rationality strengthens individual rationality since $\mu(i) \subseteq A_i \cup \Omega_i$ and $|\mu(i)| = |\Omega_i|$ imply that each object

[16] Since we assume that endowments are fixed and known, that agents only have private information about which objects they find desirable.

[17] Bogomolnaia and Moulin [2004] consider a two-sided one-to-one matching market in which each agent partitions potential match partners from the other side of the market into acceptable (better than being left unmatched) and unacceptable (worse than being left unmatched) ones. Since each agent is endowed with and cannot consume more than one "object", agents' preferences are entirely described by their sets of acceptable partners.

in $\Omega_i \setminus \mu(i)$ is replaced with an object in A_i that is at least as good under $\succsim_i^*$. Note also that responsiveness implies that agent i's ranking of two sets of objects $B_i, C_i \in X_i \cap (A_i \cup \Omega_i)$ with respect to her trichotomous preference relation $\succsim_i$ is completely determined by A_i.[18]

Aside from its normative appeal, component-wise individual rationality also allows for mechanisms that are "simple" in the sense of only having to ask agents to reveal their desirable sets. By contrast, in order to compare matchings that are not component-wise individually rational, we need agents to also provide us with information on how they trade off receiving some undesirable objects in exchange for receiving more desirable objects. Formally, let $\mathcal{A}_i = 2^O$ be the set of all possible desirable sets for agent i and let $\mathcal{A} = \times_{i \in I} \mathcal{A}_i$ be the set of all possible profiles of desirable sets. A *simple mechanism* is a mapping $\tilde{\varphi} : \mathcal{A} \to \mathcal{M}$. A simple mechanism is *component-wise individually rational*, if it always selects a component-wise individually rational matching. A simple mechanism $\tilde{\varphi}$ is *Pareto-efficient*, if, for any $\succsim \in \mathcal{T}$ with associated profile of desirable sets $A \in \mathcal{A}$, $\tilde{\varphi}(A)$ is Pareto-efficient with respect to $\succsim$. Finally, a simple mechanism $\tilde{\varphi}$ is *strategy-proof*, if no agent can ever benefit by lying about her set of desirable objects—that is, $\tilde{\varphi}$ is strategy-proof if there do not exist $\succsim \in \mathcal{T}$ with associated profile of desirable sets $A \in \mathcal{A}$, $i \in I$, and $\hat{A}_i \in \mathcal{A}_i$ such that $\tilde{\varphi}_i(\hat{A}_i, A_{-i}) \succ_i \varphi_i(A)$. We show that there is no extra cost associated with imposing the stronger requirement of component-wise individual rationality since a simple, component-wise individually rational, Pareto-efficient, and strategy-proof mechanism exists on the trichotomous domain (Theorems 1 and 3).

3 The Individually Rational Priority Algorithm

In this section, we introduce an algorithm that produces individually rational and Pareto-efficient matchings. Like serial dictatorship, agents take turns picking their most preferred sets of objects according to some exogenous priority order. The crucial difference is that our algorithm constrains each agent to choose in a way that is compatible with component-wise individual rationality for all agents. Since the set of component-wise individually rational matchings varies in response to each agent's report, an agent can influence the choice set of any other agent—not just those with lower priority. Despite this, the induced direct mechanism is strategy-proof, which we show to in the next section.

We now describe our algorithm. For the remainder of the paper, we equate an agent's index $i \in I = \{1, \dots, n\}$ with her priority, where lower indices correspond to higher priority. The counter t for the steps of the algorithm then also doubles as the reference for the successively next highest priority agent.

[18] Recall that we assume Ω to be fixed throughout.

For a general profile of preferences, $\succsim \in \mathcal{P}$—whether $\succsim \in \mathcal{T}$ or not—the *individually rational priority (IRP) algorithm* for $\succsim$ proceeds as follows:

> **Step 0**: Let $\mathcal{M}^0(\succsim)$ be the set of all individually rational matchings at $\succsim$.
>
> **Step $t \in \{1, \ldots, n\}$**: Let
>
> $$\mathcal{M}^t(\succsim) = \{\mu \in \mathcal{M}^{t-1}(\succsim) : \text{ for each } \mu' \in \mathcal{M}^{t-1}, \mu(t) \succsim_t \mu'(t)\}$$
>
> be the set of all matchings in $\mathcal{M}^{t-1}(\succsim)$ that are most preferred by t.

Though we return to this general formulation of the algorithm in Section 6, for most of our analysis, we focus on trichotomous preferences. For this domain, we define an analogue of the IRP that selects from the set of component-wise individually rational matchings. We will refer to this algorithm as the CIRP algorithm from now on. In order to formally define the algorithm and for the discussion in the rest of this section, we fix a profile of desirable sets $A \in \mathcal{A}$. Given this input, the CIRP algorithm proceeds as follows:

> **Step 0**: Let $\mathcal{M}^0(A)$ be the set of all component-wise individually rational matchings at A.
>
> **Step $t \in \{1, \ldots, n\}$**: Let
>
> $$K^t(A) = \max_{\mu \in \mathcal{M}^{t-1}(A)} |\mu(t) \cap A_t|$$
>
> be the *promise to agent t* and
>
> $$\mathcal{M}^t(A) = \{\mu \in \mathcal{M}^{t-1}(A) : |\mu(t) \cap A_t| = K^t(A)\}$$
>
> be the set of all matchings in $\mathcal{M}^{t-1}(A)$ that comply with the promise to t.[19]

At an intuitive level, the algorithm works as follows: among all component-wise individually rational matchings, let agent 1 pick those matchings that maximize the number of desirable objects she gets and collect those matchings in the set $\mathcal{M}^1(A) \subseteq \mathcal{M}^0(A)$; among all matchings in $\mathcal{M}^1(A)$, let agent 2 pick those matchings that maximize the number of desirable objects she gets and collect those matchings in the set $\mathcal{M}^2(A) \subseteq \mathcal{M}^1(A)$; ... ; among all matchings in $\mathcal{M}^{n-1}(A)$, let agent n pick those matchings that maximize the number of desirable objects she gets and collect those matchings in the set $\mathcal{M}^n(A) \subseteq \mathcal{M}^{n-1}(A)$. Note that, for any $t \leq n$, $\mathcal{M}^t(A)$ is the set of matchings that remain <u>after</u> agent t has made her pick(s).

[19] In Section 5, we show how to compute the sequences $\{K^t(A), \mathcal{M}^t(A)\}_{t=1}^n$ in polynomial time.

We now provide an example to illustrate the mechanics of CIRP.

Example 1. There are four agents $(1, 2, 3, 4)$ and six objects $(o_1, o_2, p, q, r_1, r_2)$. Endowments and desirable sets are given by the following table:

i	Ω_i	A_i
1	$\{o_1, o_2\}$	$\{o_1, p, r_2\}$
2	$\{p\}$	$\{o_1, q\}$
3	$\{q\}$	$\{o_2\}$
4	$\{r_1, r_2\}$	$\{o_2, p\}$

In the first step, we calculate the maximal number of desirable objects that agent 1 can obtain in a component-wise individually rational matching and the set of component-wise individually rational matchings that yield exactly that maximal number. It is straightforward to compute that $K^1(A) = 2$ and that $\mathcal{M}^1(A)$ consists of the following four matchings:

	1	2	3	4
μ^1	$\{o_1, p\}$	$\{q\}$	$\{o_2\}$	$\{r_1, r_2\}$
μ^2	$\{o_1, r_2\}$	$\{q\}$	$\{o_2\}$	$\{r_1, p\}$
μ^3	$\{o_1, r_2\}$	$\{p\}$	$\{q\}$	$\{r_1, o_2\}$
μ^4	$\{p, r_2\}$	$\{o_1\}$	$\{q\}$	$\{r_1, o_2\}$

Proceeding, $K^2(A) = 1$, $\mathcal{M}^2(A) = \{\mu^1, \mu^2, \mu^4\}$, $K^3(A) = 1$, and $\mathcal{M}^3(A) = \{\mu^1, \mu^2\}$. Finally, since $1 = |\mu^2(4) \cap A_4| > |\mu^1(4) \cap A_4| = 0$, we obtain that $K^4(A) = 1$ and $\mathcal{M}^4(A) = \{\mu^2\}$.

The following lemma, which follows from the description above, articulates an important property of CIRP that we use throughout our proofs.

Lemma 1. *For any pair t, t' such that $t' \leq t$ and any $\mu \in \mathcal{M}^t(A)$, we have that $|\mu(t') \cap A_{t'}| = K^{t'}(A)$.*

Lemma 1 implies, in particular, that all matchings in $\mathcal{M}^t(A)$ are *welfare equivalent* for all agents $t' \leq t$. Hence, all matchings in $\mu \in \mathcal{M}^n(A)$ are welfare equivalent for <u>all</u> agents and we refer to any matching in $\mathcal{M}^n(A)$ as a *CIRP outcome* at A.[20] The main result of this section, which we present next, shows that CIRP outcomes satisfy all of our desiderata for matchings.

Theorem 1. *Every CIRP outcome is component-wise individually rational and Pareto-efficient.*

[20]Due to indifferences in agents' preferences, $\mathcal{M}^n(A)$ need not be a singleton.

Component-wise individual rationality of CIRP outcomes follows immediately from the definition of the algorithm: For any $t \in \{1, \ldots, n\}$, we have $\mathcal{M}^t(A) \subseteq \mathcal{M}^{t-1}(A)$; hence, $\mathcal{M}^n(A) \subseteq \mathcal{M}^0(A)$, which implies that CIRP outcomes are component-wise individually rational and thus also individually rational. In the remainder of this section, we first develop a characterization of matchings in the sequence $\{\mathcal{M}^t(A)\}_{t=1}^n$ that plays a crucial role for our subsequent analysis and then use it to establish Pareto-efficiency.

Before proceeding, we observe that an important consequence of balancedness is that we can move between matchings by means of "one-to-one trading cycles" that are disjoint in terms of objects. In particular, for any t and any $\mu \in \mathcal{M}^t(A)$, μ can be obtained from Ω in this manner. We will show that matchings in $\mathcal{M}^t(A)$ can be characterized by the absence of trading cycles that lead to an increase in the number of desirable objects held by any agent $t' \leq t$. We now develop these ideas formally.

Given a matching $\mu \in \mathcal{M}^0(A)$, a *cycle of* μ is a sequence $C = (i^1, o^1, \ldots, i^M, o^M)$ of M distinct agents and M distinct objects such that, for each $m \in \{1, \ldots, M\}$, $o^{m-1} \in \mu(i^m)$ and $o^m \notin \mu(i^m)$, where $o^0 = o^M$. For the following discussion, fix a cycle $C = (i^1, o^1, \ldots, i^M, o^M)$ of μ. For any $m \in \{1, \ldots, M\}$, we say that i^m *points to* o^m and that o^m *points to* i^{m+1}, where $i^{M+1} = i^1$. Thus, C is a sequence of agent-object pairs such that each agent i^m points to an object o^m that she does not get at μ—that is, $o^m \notin \mu(i^m)$—and each object o^m points to the agent who gets it at μ—that is, $o^m \in \mu(i^{m+1})$. Let $I(C) = \{i^1, \ldots, i^M\}$ be the set of agents involved in C and let $O(C) = \{o^1, \ldots, o^M\}$ be the set of objects involved in C. Let $\mu + C$ denote the matching that results from μ by executing the trades designated by C—that is, for each $k \in I$, let

$$(\mu + C)(k) = \begin{cases} \mu(k) & \text{if } k \notin I(C) \\ (\mu(k) \setminus \{o^{m-1} : i^m = k\}) \cup \{o^m : i^m = k\} & \text{if } k \in I(C). \end{cases}$$

We say that, for any $m \in \{1, \ldots, M\}$, i^m *trades* o^{m-1} *for* o^m *in* C. Independent of preferences, based solely on the requirement of balancedness, it turns out that we can mechanically move between any pair of matchings by means of a sequence of cycles that are disjoint in terms of objects. Given a pair $\mu, \nu \in \mathcal{M}$, we say that a collection of cycles of $\{C_1, \ldots, C_K\}$ is a *decomposition of* $\nu - \mu$ if

1. $C_1, \ldots, C_K$ are disjoint in terms of objects—that is, for each pair $l, k \in \{1, \ldots, K\}$, $O(C_l) \cap O(C_k) = \emptyset$, and

2. executing all of the cycles, starting at μ, yields ν—that is,

$$(((\mu + C_1) + C_2) + \ldots) + C_K = \nu.$$[21]

The next lemma states that such a decomposition exists for any pair of matchings.[22]

Lemma 2. *For each pair $\mu, \nu \in \mathcal{M}$, there is a decomposition of $\nu - \mu$.*

We now define some additional properties of cycles based on agents' preferences. We say that C

- is *component-wise individually rational* if, for each $m \in \{1, \ldots, M\}$, $o^m \in \Omega_{i^m} \cup A_{i^m}$,
- *increases i's welfare* if $|(\mu + C)(i) \cap A_i| > |\mu(i) \cap A_i|$,
- *decreases i's welfare* if $|(\mu + C)(i) \cap A_i| < |\mu(i) \cap A_i|$, and
- *affects i* if it either increases or decreases i's welfare.
- *Pareto-improving* if it increases some agent's welfare without decreasing that of any other agent.

Since preferences are trichotomous and exchange is balanced, it is intuitive that a complex trade, that is beneficial to all agents involved, can be broken down into many simpler trades involving just one object per agent. The next result formalizes this intuition showing that any Pareto-improvement from a component-wise individually rational matching can be achieved by several separate such cycles.

Proposition 1. *If μ is component-wise individually rational at A, then μ is Pareto-efficient if and only if there is no component-wise individually rational and Pareto-improving cycle of μ.*

Before we complete our proof of Theorem 1, we introduce one more useful concept. We say that C is a *constrained improvement cycle (CIC) of μ for t at A* if C

1. is component-wise individually rational,
2. increases t's welfare, and
3. does not affect any agent $t' < t$.

Finally, we say that C is a *CIC of μ at A*, if there exists a t such that C is a CIC of μ for t at A. Note that if C is component-wise individually rational and Pareto-improving, then C is a CIC of μ at A. Now, with the above definitions in hand, we have the following.

[21] Since the cycles are disjoint in terms of objects, the order they are executed in does not matter.

[22] There could be more than one such decomposition.

Theorem 2. *For any $t \in \{1, \dots, n\}$, $\mu \in \mathcal{M}^t(A)$ if and only if $\mu \in \mathcal{M}^{t-1}(A)$ and there does not exist a CIC of μ for t at A.*

Theorem 2 and Proposition 1 imply Theorem 1: by Theorem 2, there cannot be a CIC of any CIRP outcome; by Proposition 1, a matching is Pareto-efficient if it does not admit a CIC.

To gain some intuition for Theorem 2, fix some $t \leq n$ and consider the t^{th} step of the IRP algorithm. Here, the number of desirable objects awarded to agent t is determined subject to the constraints imposed by the promises to agents $t' < t$. These constraints are expressed in the restriction to matchings in $\mathcal{M}^{t-1}(A)$. Now fix a matching $\mu \in \mathcal{M}^{t-1}(A)$. For the necessity part of Theorem 2, assume that there is a CIC C of μ for t at A. Since C is a CIC for t, $\mu + C$ keeps all promises to agents $t' < t$ and $\mu + C$ is individually rational. Given that C increases the welfare of t, we find that $\mu \notin \mathcal{M}^t(A)$. For the sufficiency part of Theorem 2, assume that $\mu \notin \mathcal{M}^t(A)$ and let $\nu \in \mathcal{M}^{t-1}(A)$ be an arbitrary matching that leaves t strictly better off than μ. Then there must exist an object o that is desirable for t and that t gets in ν but not in μ. In our proof, we use o as well as the other assignments in μ and ν to construct a CIC of μ for j at A.

4 CIRP Mechanisms and Strategy-proofness

Though we have fixed the priority over the agents, there are multiple direct mechanisms that are induced by the CIRP algorithm: every selection φ from $\mathcal{M}^n$. By Lemma 1, they are welfare equivalent. We call any such mechanism a *component-wise individually rational priority (CIRP) mechanism*. In this section, we establish that any CIRP mechanism is strategy-proof.

For the remainder of this section, fix a true profile of trichotomous preferences $\succsim \in \mathcal{T}$ with corresponding true profile of desirable sets $A \in \mathcal{A}$, an agent $i \in I$, and a possible manipulation for i, $\hat{A}_i \in \mathcal{A}_i$. Denote the profile $(\hat{A}_i, A_{-i})$ by $\hat{A}$. Recall that $\{\mathcal{M}^t(A)\}_{t=1}^n$ and $\{\mathcal{M}^t(\hat{A})\}_{t=1}^n$ summarize the progression of CIRP for inputs A and $\hat{A}$ respectively.

We start by showing that i cannot increase her number of truly desirable objects by falsely claiming that an object is desirable when it is not.

Lemma 3. *If, for some $o \notin A_i$, $\hat{A}_i = A_i \cup \{o\}$, then, for each $\mu \in \mathcal{M}^i(\hat{A})$, $|\mu(i) \cap A_i| \leq K^i(A)$.*

Next, we consider the case where i falsely declares a desirable object to be undesirable.

Lemma 4. *If, for some $o \in A_i$, $\hat{A}_i = A_i \setminus \{o\}$, then $K^i(A) \geq K^i(\hat{A})$.*

These two lemmas are two sides of the same coin. The reasoning behind them, like Proposition 1, is that when preferences are trichotomous and exchange is balanced, complex trades can be broken down into simpler one-for-one trades. Indeed, any matching $\mu \in \mathcal{M}^n(A)$ is the result of applying *some* set of cycles, which describe one-for-one trades, to the endowment. If an object $o \in A_i$ is removed to i's desirable set, and it changes the number of desirable objects that i gets, then it has to be that i was getting o in μ. In particular, i had to have gotten o in one of the cycles taking us from the endowment to μ. Showing that only that cycle is affected by the removal of o from i's desirable set is the heart of the proof of these two lemmas.

Finally, we appeal to Lemmas 3 and 4 to prove that any CIRP mechanism is strategy-proof.

Theorem 3. *Any CIRP mechanism is strategy-proof.*

We point out that our restriction to trichotomous preferences is important for strategy-proofness of CIRP mechanisms. For example, even for the case of singleton endowments, a sequential priority over the set of individually rational matchings is not generally strategy-proof if agents' preferences have more than three indifference classes (see Example 3).

5 Computing CIRP Matchings

In this section, we define a polynomial time algorithm (Algorithm 1) to compute matchings in $\mathcal{M}^n(A)$ based on network flows. The procedure described in Section 3 enumerates of all of the individually rational matchings ($\mathcal{M}^0(A)$), the subset of these matchings in which 1 receives the most desirable objects ($\mathcal{M}^1(A)$), and so on. The problem is that the sizes of these sets are an exponential function of the number of agents and objects. The algorithm that we present here uses the fact that each of these sets is the set of solutions to a particular network flow problem.

For the following discussion, fix an endowment Ω and a profile of desirable sets A. We work with a directed graph (V, E). The set of vertices V contains three vertices corresponding to each $i \in I$, call them i, i^A, and i^U. It also contains one vertex corresponding to each $o \in O$. Finally, it contains a source S and a sink T. That is,

$$V = \{i, i^A, i^U : i \in I\} \cup O \cup \{S, T\}$$

The set of edges E contains an edge from S to each $i \in I$, from each $i \in I$ to i^A and i^U, from

i^A to each object in A_i, from i^U to each object in $\Omega_i \setminus A_i$, and from each $o \in O$ to T. That is,

$$
\begin{aligned}
E &= \{(S,i) : i \in I\} \\
&\cup \{(i,i^A),(i,i^U) : i \in I\} \\
&\cup \{(i^A,o) : o \in A_i\} \\
&\cup \{(i^U,o) : o \in \Omega_i \setminus A_i\} \\
&\cup \{(o,T) : o \in O\}.
\end{aligned}
$$

Roughly speaking, copy i^A of agent i collects all desirable objects, while copy i^U collects all endowments that are not desirable. Throughout our algorithm, we only vary the capacities of each edge. The capacity of edge $e \in E$ is $q(e)$. We initialize q as follows:

$$
\begin{array}{rcll}
q(S,i) &=& |\Omega_i| & i \in I \\
q(i,i^A) &=& |\Omega_i| & i \in I \\
q(i,i^U) &=& |\Omega_i \setminus A_i| & i \in I \\
q(i^A,o) &=& 1 & i \in I, o \in A_i \\
q(i^U,o) &=& 1 & i \in I, o \in \Omega_i \setminus A_i \\
q(o,T) &=& 1 & o \in O.
\end{array}
$$

We say that $f : E \to \mathbb{N}$ is an (integer) flow, if $f(e) \le q(e)$, for all $e \in E$, and $\sum_{e \in \delta^-(v)} f(e) = \sum_{e \in \delta^+(v)} f(e)$, for all $v \in V \setminus \{S,T\}$.[23] The value of a flow f through i is $f(S,i)$ and the value of a flow f is $v(f) = \sum_{i \in I} f(S,i)$. Since $q(S,i) = |\Omega_i|$, for all $i \in I$, the maximum flow through i cannot be larger than $|\Omega_i|$. Since, for all $o \in O$, $i \in I$, and $D \in \{A,U\}$, $q(i^D,o) \in \{0,1\}$ and $q(o,T) = 1$, a flow cannot use two edges involving the same object but different agents. If no flow g has higher value than f, then f is a *maximum flow*. Let $\textsc{MaxFlow}(q)$ be the value of a maximum flow from S to T when the capacities are q. By the above observation that the flow through i cannot exceed $|\Omega_i|$, for every q that we consider, $\textsc{MaxFlow}(q) \le \sum_{i \in I} |\Omega_i|$ and, for the initial value of q, $\textsc{MaxFlow}(q) = \sum_{i \in I} |\Omega_i|$ since we can assign each agent her endowment at that capacity vector. Given $\mu \in \mathcal{M}^0(A)$, we say that f *corresponds to* μ if, for each $i \in I$ and each $o \in \mu(i) \cap A_i$, $f(i^A,o) = 1$ and, for each $i \in I$ and each $o \in \mu(i) \cap \Omega_i \setminus A_i$, $f(i^u,o) = 1$.

Algorithm 1 works by setting up an initial network such that the set of all maximum flows corresponds to the set of all individually rational matchings—that is, $\mathcal{M}^0(A)$. In general, at the start of the t^{th} iteration, the network is such that the set of all maximum flows corresponds to $\mathcal{M}^{t-1}(A)$. The algorithm then modifies the network by repeatedly decreasing t's *capacity for endowments that are not desirable*, $q(t,t^U)$, as long as the maximum flow equals $\sum_i |\Omega_i|$.

[23] $\delta^+(v)$ is the set of edges that originate at v and $\delta^-(v)$ is the set of edges that end at v.

This ensures that the maximum flows correspond to some matchings. In particular, they correspond to a subset of $\mathcal{M}^{t-1}(A)$. The flow along (t, t^U) represents the number of objects in $\Omega_t \setminus A_t$ that t gets. So if this number cannot be decreased further, then we have maximized the number of objects in A_t that t gets. Thus, at the end of the t^{th} iteration, the set of all maximum flows in the network corresponds to $\mathcal{M}^t(A)$. Thus, when the algorithm terminates, the set of all maximum flows in the network corresponds to $\mathcal{M}^n(A)$.

Algorithm 1 Procedure to compute CIRP

procedure CIRP(V, E, q)
    for $t = 1$ to $t = N$ **do** ▷ Loop invariant: $\textsc{MaxFlow}(q) = \sum_i |\Omega_i|$
        if $q(t, t^U) > 0$ **then** ▷ t has capacity
            $\hat{q} = q$
            $\hat{q}(t, t^U) = \hat{q}(t, t^U) - 1$ ▷ Decrement t's capacity by 1
            while $\textsc{MaxFlow}(\hat{q}) = \sum_i |\Omega_i|$ **do** ▷ As long as the loop invariant holds
                $q = \hat{q}$
                $\hat{q}(t, t^U) = \hat{q}(t, t^U) - 1$ ▷ Decrement t's capacity by 1
    return q

The following result summarizes the key properties of Algorithm 1.

Theorem 4. *Let $A \in \mathcal{A}^n$ be a profile of desirable sets and let Ω be an endowment. Suppose that V, E, and q are defined based on A and Ω as described above.*

1. *A flow f is a maximum flow in $(V, E, \mathrm{CIRP}(V, E, q))$ if and only if it corresponds to a CIRP outcome.*

2. *The complexity of Algorithm 1 is $O(mn^3 \log(k))$, where $m = |O|$ and $k = \max_{i \in I} |\Omega_i|$.*

6 Beyond Trichotomous Preferences

The remarkable aspect of our results is that the very simple approach of CIRP yields component-wise individually rational, Pareto-efficient, and strategy-proof mechanisms. Our positive results, of course, rely on the assumption that agents' preferences are trichotomous. Consequently, we view CIRP as a potentially useful tool for the exchange of bundles of indivisible goods like work shifts. In this section, we discuss whether the CIRP-approach can be fruitful beyond trichotomous preferences. The exercise of characterizing a maximal domain for the existence of mechanisms that satisfy our three desiderata is beyond the scope of this paper. Nonetheless, we demonstrate some ways in which we can and cannot relax our assumptions on preferences while maintaining the desirable properties of CIRP. We first discuss more general responsive domains and then shift attention to complementarities in agents' preferences. Since we consider non-trichotomous domains, we rely on the more general formulation of IRP in Section 3.

6.1 More on responsive preferences

In this subsection, we provide a series of examples which suggest that it is very hard to move beyond the trichotomous domain if one wants to maintain the assumption of responsive preferences. Specifically, in order for IRP to be strategy-proof, it is essential that each agent partitions other agents' endowed objects into desirable/undesirable and ranks others' desirable objects at least as highly as any of her own endowed objects. Our first example shows that CIRP is not strategy-proof if an agent can rank another agent's endowed object strictly between two of her own endowed objects.

Example 2. There are three agents $(1, 2, 3)$ and four objects $(\omega_1, \omega_2, \omega_3, \omega_3')$. Endowments are given by $\Omega_1 = \{\omega_1\}$, $\Omega_2 = \{\omega_2\}$, and $\Omega_3 = \{\omega_3, \omega_3'\}$. Agents' true preferences are trichotomous with desirable sets $A_1 = \{\omega_2, \omega_3\}$, $A_2 = \{\omega_3'\}$, and $A_3 = \{\omega_1, \omega_3'\}$. For this input, IRP produces a matching in which 3 ends up with $\{\omega_1, \omega_3\}$. Now assume that 3 can have any responsive ordering $\succsim_3$ with the following ordering over individual objects: $\omega_3' \succ_3^* \omega_1 \succ_3^* \omega_3$. One possibility for such an ordering is that $\{\omega_3, \omega_3'\} \succ_3 \{\omega_1, \omega_3\}$. If the preferences of agents 1 and 2 are left unchanged, the unique IRP outcome now assigns the bundle $\{\omega_1, \omega_3'\}$ to agent 3.

Next, we consider the case where an agent might discriminate between two distinct objects initially owned by others.

Example 3. There are three agents $(1, 2, 3)$ and three objects $(\omega_1, \omega_2, \omega_3)$. Endowments are given by $\Omega_1 = \{\omega_1\}$, $\Omega_2 = \{\omega_2\}$, and $\Omega_3 = \{\omega_3\}$. Agents 1 and 2 have trichotomous preferences with desirable sets $A_1 = A_2 = \{\omega_3\}$. Agent 3's true preferences are responsive with respect to the following ranking of individual objects: $\omega_2 \succ_3^* \omega_1 \succsim_3^* \omega_3$. For this input, IRP produces a matching in which 3 ends up with ω_1. Suppose 3 were to deviate to reporting preferences that are responsive with respect to the ordering: $\omega_2 \tilde{\succ}_3^* \omega_3 \tilde{\succ}_3^* \omega_1$.[24] If the preferences of agents 1 and 2 are left unchanged, the unique IRP outcome assigns ω_2 to 3.

6.2 Domains with Complementarities

So far, we have only considered responsive preferences. Two key features of the responsive domain are that we can evaluate a particular trade without reference to other trades and

[24]Such an ordering is compatible with the assumption of trichotomous preferences.

that an agent finds it worthwhile to engage in exchange as long as there is a one-for-one trade that makes her better off. The literature on matching theory and its applications is rich with examples in which neither of the two features are present. Here, we consider domains with complementarities.

In our first example one agent views two objects as perfect complements while maintaining the other features of our base model.

Example 4. There are three agents $(1, 2, 3)$ and six objects $(\omega_1, \omega_1', \omega_2, \omega_2', \omega_3, \omega_3')$. Endowments are given by $\Omega_1 = \{\omega_1, \omega_1'\}$, $\Omega_2 = \{\omega_2, \omega_2'\}$, and $\Omega_3 = \{\omega_3, \omega_3'\}$. Agents 1 and 3 have trichotomous preferences with desirable sets $A_1 = \{\omega_3, \omega_3', \omega_2\}$ and $A_3 = \{\omega_1, \omega_1', \omega_2'\}$. Moreover, agent 1 prefers her endowment to any set containing ω_2'—that is, $\{\omega_1, \omega_1'\} \succ_1 B$ for all B such that $\omega_2' \in B$. Agent 2, on the other hand, is only willing to trade if she can get both ω_1 and ω_1'. Put differently, the only two sets of objects that satisfy the individual rationality constraint of agent 2 are $\{\omega_1, \omega_1'\}$ and $\{\omega_2, \omega_2'\}$.[25] For this input, IRP produces a matching in which 2 ends up with $\{\omega_1, \omega_1'\}$. Hence, 3 ends up with ω_2' and either ω_3 or ω_3'. However, if 3 deviates to reporting a desirable set consisting only of $\{\omega_1, \omega_1'\}$, IRP will match 2 to exactly these two objects.

We now show that IRP is a fruitful approach to problems in which agents are willing to participate in exchange if and only if they end up with a specific *target number* of desirable objects. We first define such preferences formally:

Definition 2 (All-or-nothing with respect to A_i, Ω_i, and k_i)**.** Fix an agent $i \in I$, an endowment $\Omega_i \subseteq O$, a desirable set A_i, and a target $k_i \in \mathbb{Z}_+$. We say that $\succsim_i$ satisfies *all-or-nothing* with respect to A_i, Ω_i and k_i if,

1. for any $B_i \in X_i$, $B_i \sim_i \Omega_i$ if and only if $B_i = \Omega_i$,
2. for any $B_i \in X_i$, $B_i \succ_i \Omega_i$ if and only if both $B_i \subseteq A_i \cup \Omega_i$ and $B_i \cap A_i = k_i$, and
3. for any pair $B_i, C_i \in X_i$ such that $B_i, C_i \succ_i \Omega_i$, $B_i \sim_i C_i$.

"All-or-nothing"-preferences are related to our trichotomous domain in the sense that we require each agent to partition objects into desirable and undesirable ones. A significant departure is that an agent's welfare is not strictly monotone in the number of desirable objects she obtains.[26] Rather, agent i derives utility from exchange if and only if she exactly

[25] Another possible interpretation of the example is that 2 has trichotomous preferences with respect to the desirable set $\{\omega_1, \omega_1'\}$ but is only willing to trade if she ends up with at least two desirable objects. This could be, for instance, due to the presence of fixed costs for participating in the exchange.

[26] Note that the preferences in Example 4 do not satisfy the just described assumption since the welfare of agents 1 and 3 is strictly increasing in the number of desirable objects that they receive.

hits her target. While "all-or-noting" may not be the most reasonable assumption in the context of shift exchange, it is a better model for other potential applications. For example, following Ergin et al. [2017], a recent literature has started investigating the design of organ exchanges for applications in which patients may require more than one donor.[27] If patients need multiple donors in order to be treated successfully, it is reasonable to expect that they (and their incompatible direct donors) are only willing to engage in exchange if they end up with exactly the required number of compatible donors. Our next result is that IRP satisfies all of our desiderata for "all-or-nothing"-preferences.

Proposition 2. *For "all-or-nothing"-preferences, any IRP mechanism is individually rational, Pareto-efficient, and strategy-proof.*

7 Conclusion

Without restrictions on the preference domain, the three central requirements of individual rationality, Pareto-efficiency, and strategy-proofness are incompatible for the exchange of indivisible goods with multi-unit demand and supply in the absence of monetary transfers. In this paper, we have focused on balanced exchange and identified a meaningful restriction on preferences where Pareto-efficiency and strategy-proofness are actually compatible with a strong version of individual rationality when endowments are known.

Our results for this model show that, despite earlier negative results, this is an area ripe for future research.[28] We conclude with some possible directions for future research.

The IRP approach is simple and intuitive. Though individual rationality, Pareto-efficiency, and strategy-proofness are incompatible over the unrestricted domain, we have shown that our approach does extend to some non-trichotomous domains. A systematic study of such domains would be an interesting avenue of study. Moreover, delineating maximal domains where our three axioms are compatible is also an important, albeit difficult, question. Despite the desirable normative and strategic properties of IRP mechanisms for all-or-nothing preferences, it is not clear that IRP outcomes can be computed efficiently for this domain. Again, the efficient computation of individually rational and Pareto-efficient matchings beyond trichotomous preferences remains an open question. Since our analysis is for fixed endowments, a systematic analysis of agents' incentives to reveal their endowments remains another interesting area for further research. As we have demonstrated in Section 6,

[27] Another interesting complication (which is well beyond the scope of our analysis) in these problems is that there may be several different ways in which a given donor may donate to a given patient, some of them associated with more risk for donors than others. Ergin et al. [2020] show how to design liver exchange mechanisms that elicit patient-donor pairs' preferences over different donation options truthfully.

[28] See also the concurrent and independent positive results of Andersson et al. [forthcoming].

overcoming the incompatibility of Pareto-efficiency and strategy-proofness among individually rational mechanisms requires rather strong restrictions on preferences. An alternative in the literature to restricting preferences has been to compromise on Pareto-efficiency. One might instead consider a mechanism that cannot be Pareto-improved upon, at every preference profile, by any other individually rational and strategy-proof mechanism to be "constrained efficient."[29] Our results suggest that there is a tradeoff between the inefficiency of such a constrained efficient mechanism and the permissibility of the preference domain since full efficiency is possible on the trichotomous domain.

References

M. Akbarpour, S. Li, and S. O. Gharan. Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783 – 815, March 2020.

J. Alcalde-Unzu and E. Molis. Exchange of indivisible goods and indifferences: The top trading absorbing sets mechanisms. *Games and Economic Behavior*, 73(1):1–16, September 2011.

T. Andersson, Ágnes Cseh, L. Ehlers, and A. Erlanson. Organizing time banks: Lessons from matching markets. *American Economic Journal:Microeconomics*, forthcoming.

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

M. Atlamaz and B. Klaus. Manipulation via Endowments in Exchange Markets with Indivisible Goods. *Social Choice and Welfare*, 28(1):1–18, January 2007.

H. Aziz and B. De Keijzer. Housing markets with indifferences: A tale of two mechanisms. In *AAAI*, volume 12, pages 1–1, 2012.

P. Biró, F. Klijn, and S. Pàpai. Balanced exchange in a multi-object shapley-scarf market. Working paper, 2018.

A. Bogomolnaia and H. Moulin. Random matching under dichotomous preferences. *Econometrica*, 72(1):257–279, 01 2004.

[29]See Klaus [2008] and Anno and Kurino [2016] for instances of this approach in the multi-type object allocation model.

E. Budish and J. B. Kessler. Bringing Real Market Participants' Real Preferences into the Lab: An Experiment that Changed the Course Allocation Mechanism at Wharton. NBER Working Papers 22448, National Bureau of Economic Research, Inc, July 2018.

A. Chernev. When more is less and less is more: The role of ideal point availability and assortment in consumer choice. *Journal of Consumer Research*, 30(2):170–183, 2003.

R. Dhar. Consumer preference for a no-choice option. *Journal of Consumer Research*, 25(2): 215–231, 1997.

U. M. Dur and M. U. Ünver. Two-sided matching via balanced exchange. *Journal of Political Economy*, 123(5):1156 – 1177, June 2019.

H. Ergin, T. Sönmez, and M. U. Ünver. Dual-donor organ exchange. *Econometrica*, 85(5): 1645 – 1671, September 2017.

H. Ergin, T. Sönmez, and M. U. Ünver. Efficient and incentive-compatible liver exchange. *Econometrica*, 88(3):1645 – 1671, May 2020.

E. Fujita, J. Lesca, A. Sonoda, T. Todo, and M. Yokoo. A complexity approach for core-selecting exchange with multiple indivisible goods under lexicographic preferences. In *Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence*, AAAI'15, pages 907–913, 2015. ISBN 0-262-51129-0.

E. A. Greenleaf and D. R. Lehmann. Reasons for substantial delay in consumer decision making. *Journal of Consumer Research*, 22(2):186–199, 1995.

J. W. Hatfield. Pairwise kidney exchange: Comment. *Journal of Economic Theory*, 125(2): 189–193, 2005.

S. S. Iyengar and M. R. Lepper. When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology*, 79(6):995–1006, 2000.

P. Jaramillo and V. Manjunath. The difference indifference makes in strategy-proof allocation of objects. *Journal of Economic Theory*, 147(5):1913–1946, 2012.

B. Klaus. The coordinate-wise core for multiple-type housing markets is second-best incentive compatible. *Journal of Mathematical Economics*, 44(9):919 – 924, 2008.

B. Klaus, D. Dimitrov, and C.-J. Haake. Bundling in exchange markets with indivisible goods. *Economics Letters*, 93(1):106–110, 2006.

H. Konishi, T. Quint, and J. Wako. On the shapley,Äìscarf economy: the case of multiple types of indivisible goods. *Journal of Mathematical Economics*, 35(1):1 – 15, 2001. ISSN 0304-4068.

P. Milgrom. Assignment Messages and Exchanges. *American Economic Journal: Microeconomics*, 1(2):95–113, August 2009.

P. Milgrom. Simplified mechanisms with an application to sponsored-search auctions. *Games and Economic Behavior*, 70(1):62–70, September 2010.

P. Milgrom. Critical issues in the practice of market design. *Economic Inquiry*, 49(2):311–320, 2011.

D. Monte and N. Tumennasan. Centralized allocation in multiple markets. *Journal of Mathematical Economics*, 61:74 – 85, 2015.

H. Moulin. *Cooperative Microeconomics: A Game-Theoretic Introduction.* Princeton University Press, 1995.

A. Othman, T. Sandholm, and E. Budish. Finding approximate competitive equilibria: Efficient and fair course allocation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: Volume 1 - Volume 1*, AAMAS '10, page 873–880. International Foundation for Autonomous Agents and Multiagent Systems, 2010. ISBN 9780982657119.

E. Perez-Richet. A note on the tight simplification of mechanisms. *Economics Letters*, 110 (1):15 – 17, 2011.

W. Phan and C. Purcell. On the complexity of manipulation in the exchange of indivisible objects. Working paper, 2018.

M. S. Pini, F. Rossi, K. B. Venable, and T. Walsh. Manipulation complexity and gender neutrality in stable marriage procedures. *Autonomous Agents and Multi-Agent Systems*, 22 (1):183–199, Jan 2011.

A. E. Roth. Incentive compatibility in a market with indivisible goods. *Economics Letters*, 9 (2):127–132, 1982.

A. E. Roth and A. Postlewaite. Weak versus strong domination in a market with indivisible goods. *Journal of Mathematical Economics*, 4(2):131–137, August 1977.

A. E. Roth, T. Sönmez, and M. U. Ünver. Pairwise kidney exchange. *Journal of Economic Theory*, 125(2):151 – 188, 2005. ISSN 0022-0531.

D. Saban and J. Sethuraman. House allocation with indifferences: A generalization and a unified view. In *Proceedings of the Fourteenth ACM Conference on Electronic Commerce*, EC '13, pages 803–820, 2013. ISBN 978-1-4503-1962-1.

L. Shapley and H. Scarf. On cores and indivisibility. *Journal of Mathematical Economics*, 1: 23–37, 1974.

S. Sikdar, S. Adali, and L. Xia. Mechanism design for multi-type housing markets. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, AAAI'17, pages 684–690, 2017.

S. Sikdar, S. Adali, and L. Xia. Top-trading-cycles mechanisms with acceptable bundles. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence*, AAAI'18, 2018.

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

T. Sönmez and M. U. Ünver. Altruistically unbalanced kidney exchange. *Journal of Economic Theory*, 152(C):105–129, 2014.

Appendices

A Proofs from Section 3

Proof of Lemma 2. We start by defining the distance between a pair of matchings. Consider a pair of matchings $\mu^1, \mu^2 \in \mathcal{M}$. For each agent $i \in I$, $|\mu^1(i) \cap \mu^2(i)|$ is the number of objects in common between the two matchings, and, since i consumes Ω_i objects, $|\Omega_i| - |\mu^1(i) \cap \mu^2(i)|$ is the number objects that differ between the two matchings.[30] The *distance* between the two is simply the sum of this number across all agents:

$$d(\mu^1, \mu^2) = \sum_{i \in I} |\Omega_i| - |\mu^1(i) \cap \mu^2(i)|.$$

[30] To be more precise, $|\Omega_i| - |\mu^1(i) \cap \mu^2(i)|$ is the number of slots in i's "schedule" that are filled differently. The number of objects that differ between the two matchings is exactly twice the number of slots that are filled differently.

For the remainder of this proof, fix some matching μ. We proceed by induction on $d(\mu,\nu)$. As an induction base, we note that if $d(\mu,\nu) = 0$, then $\mu = \nu$ and there is nothing left to show. Suppose that there exists some $K > 0$ such that for every ν' which satisfies $d(\mu,\nu') < K$ there is a decomposition of $\nu' - \mu$. Let ν be a matching such that $d(\mu,\nu) = K$.

Consider a directed graph on the vertices

$$\{i : i \in I \text{ and } \mu(i) \neq \nu(i)\} \cup \{o : \text{there are } i, j \in I \text{ such that } i \neq j \text{ and } o \in \mu(i) \cap \nu(j)\}.$$

So there is a vertex for each agent who receives a different set from the two matchings and for each object assigned to different agents by the two matchings. Since $\mu \neq \nu$, this set of vertices is not empty. Let the set of edges be

$$\{(i,o) : i \in I \text{ and } o \in \mu(i) \setminus \nu(i)\} \cup \{(o,i) : i \in I \text{ and } o \in \nu(i) \setminus \mu(i)\}.$$

So, for each agent, there is an edge *to* every object that she receives from μ (but not from ν) and an edge *from* every object that she receives from ν (but not from μ). Every vertex in this finite graph has out-degree of at least one, so this graph has a cycle.[31] Let C be such a cycle. Since this graph is bipartite, C contains an even number of vertices, alternating between vertices in I and in O. Moreover, by definition of the set of edges, there are no cycles that contain less than two agents or less than two objects, so there are at least two elements of I in C. Thus, C is of the form $(i^1, o^1, \dots, i^M, o^M)$, where $M \geq 2$ and, for each $k \in \{1, \dots, M\}, o^{k-1} \in \nu(i^k) \setminus \mu(i^k)$ and $o^k \in \mu(i^k) \setminus \nu(i^k)$.

In the remainder of the proof, we will use C and the inductive assumption to find the desired decomposition of $\mu - \nu$. Note first that C is a cycle of ν that decreases the distance to μ. Hence, letting $\nu' = \nu + C$, we obtain a matching such that $d(\mu,\nu') < K$. By the inductive assumption, there is a decomposition $\{C_1, \dots, C_K\}$ of $\nu' - \mu$. Next, let

$$C_{K+1} = (i^1, o^M, \dots, i^2, o^1).$$

be the cycle that reverses all trades in C. Since $\nu + C = \nu'$, we have that $\nu' + C_{K+1} = \nu$. Thus, $(((\mu + C_1) + \dots) + C_K) + C_{K+1} = \nu$. Moreover, since, for each k, $o^k \in \nu'(i^k)$, we have that $o^k \notin O(C_1) \cup \dots \cup O(C_K)$. Thus, $\{C_1, \dots, C_{K+1}\}$ is a decomposition of $\nu - \mu$. □

Proof of Proposition 1. The "only if" part is trivial so we only establish the "if" part. Fix a component-wise individually rational matching μ at A. Assume that there is a matching ν that Pareto-dominates μ. By Lemma 2, there is a decomposition $\{C_1, \dots, C_K\}$ of $\nu - \mu$. We

[31] Otherwise, a depth first search would continue infinitely, since there are no terminal vertices, contradicting the finiteness of the graph.

say that C' is a subcycle of $\nu - \mu$, if there exists a decomposition $\{C_1, \ldots, C_K\}$ of $\nu - \mu$ such that $C_k = C'$ for some k.[32] If $C' = (i^1, o^1, \ldots, i^M, o^M)$ is a subcycle of $\nu - \mu$, we let

$$\nu - C' = \nu + (i^1, o^M, i^M, o^{M-1}, \ldots, i^2, o^1).$$

Note that $\nu - C'$ reverses the trades in C' so that the associated objects return into the possession of those agents who owned them at μ. Finally, we say that ν is a minimal Pareto-improvement of μ if there is no non-empty subcycle C' of $\nu - \mu$ such that $\nu - C'$ Pareto-improves upon μ.

To complete the proof, we establish that it is without loss of generality to assume that there is a cycle C such that $\mu + C = \nu$. First, it is without loss to assume that ν is a minimal Pareto-improvement. Next, we use μ and ν to construct a directed graph. Let

$$I' = \{i : i \in I \text{ and } \mu(i) \neq \nu(i)\}$$

and

$$O' = \{o : \text{there are } i, j \in I' \text{ such that } i \neq j \text{ and } o \in \mu(i) \cap \nu(j)\},$$

and let the set of vertices be $I' \cup O'$. For each $o \in O'$ and $i \in I'$ such that $o \in \mu(i) \setminus \nu(i)$, insert the edge (o, i). For each $i \in I'$, let $o^i \in O'$ be one of i's most preferred objects in $\nu(i) \setminus \mu(i)$—that is, $o^i \in \nu(i) \setminus \mu(i)$ and, for each $o' \in \nu(i) \setminus \mu(i)$, $o^i \succsim_i^* o'$— and insert the edge (i, o^i). Let G denote the resulting directed graph.

Each vertex in G has a single out edge and there is no edge from any $i \in I'$ to $o \in \mu(i)$. Therefore, G has some cycle $C = (j^1, p^1, \ldots, j^K, p^K)$ that involves at least two agents.

We now show that C weakly increases the welfare of all agents, who are involved in it. For any k such that $p^k \in A_{j^k}$, C weakly increases the welfare of j^k. Now consider some k such that $p^k \in O \setminus A_{j^k}$. Since p^k is one of j^k's most preferred objects in $\nu(j^k) \setminus \mu(j^k)$, we must have $\nu(j^k) \setminus \mu(j^k) \subseteq O \setminus A_{j^k}$. Since $\nu(j^k) \succsim \mu(j^k)$ and since agent j^k's preferences are trichotomous, we must have $\mu(j^k) \setminus \nu(j^k) \subseteq O \setminus A_{j^k}$ as well. Furthermore, given that μ is component-wise individually rational, we obtain $\mu(j^k) \setminus \nu(j^k) \subseteq \Omega_{j^k} \setminus A_{j^k}$ and $\nu(j^k) \setminus \mu(j^k) \subseteq \Omega_{j^k} \setminus A_{j^k}$. Hence, C must leave j^k unaffected.

If C is not a Pareto-improving cycle of μ, then it does not affect any agent. However, this means $\nu - C$ is a Pareto-improvement of μ, contradicting the minimality of ν. Hence, C is Pareto-improving and by the minimality of ν, $\nu = \mu + C$. □

Proof of Theorem 2. Let μ be a component-wise individually rational matching. We start our proof by verifying that if C is a component-wise individually rational cycle, then $\mu + C$ is

[32] There may be more than one decomposition of $\nu - \mu$.

component-wise individually rational as well.

Lemma 5. *Fix a profile $A \in \mathcal{A}$ and some $\mu \in \mathcal{M}^0(A)$. If C is a component-wise individually rational cycle of μ at A, then $\mu + C \in \mathcal{M}^0(A)$.*

Proof. Let $C = (i^1, o^1, \dots, i^M, o^M)$ and $\nu = \mu + C$. First, for each $o \in O$, since the objects in C are distinct, o cannot be in both $\nu(k)$ and $\nu(l)$ for distinct agents k and l. Next, for each $k \in I$, $\nu(k) \subseteq A_k \cup \Omega_k$ since $\mu(k) \subseteq A_k \cup \Omega_k$ and $\nu(k) \setminus \mu(k) \subseteq A_k \cup \Omega_k$ given that C is component-wise individually rational. Finally, we verify that for each $k \in I$, $|\nu(k)| = |\Omega_k|$. Since the objects in C are distinct and since $o^{m-1} \in \mu(i^m)$ as well as $o^m \notin \mu(i^m)$ for all m, we have

$$\begin{aligned} |\nu(k)| &= |(\mu(k) \setminus \{o^{m-1} : i^m = k\}) \cup \{o^m : i^m = k\}| \\ &= |\mu(k)| - |\{o^{m-1} : i^m = k\}| + |\{o^m : i^m = k\}| \\ &= |\mu(k)| \\ &= |\Omega_k| \end{aligned}$$

Since k was arbitrary, the last observation completes the proof. □

For the remainder of the proof, we fix $t \in \{1, \dots, n\}$. We first establish the necessity part of Theorem 2 and show that the absence of CICs is necessary for a matching to be compatible with the guarantee for t at A.

Lemma 6. *If $\mu \in \mathcal{M}^t(A)$, then there is no CIC of μ for any $t' \leq t$ at A.*

Proof. Let $\mu \in \mathcal{M}^t(A)$. Suppose to the contrary that $C = (i^1, o^1, \dots, i^M, o^M)$ is a CIC of μ for some $t' \leq t$ at A. Let $\nu = \mu + C$. By Lemma 5, ν is component-wise individually rational—that is, $\nu \in \mathcal{M}^0(A)$.

We first argue that $\nu \in \mathcal{M}^{t'-1}(A)$. If $t' = 1$, this follows by the definition of $\mathcal{M}^0(A)$ as the set of all component-wise individually rational matchings. Suppose $t' > 1$ and $\nu \in \mathcal{M}^{t''}(A)$ for some $t'' < t' - 1$. Since C does not affect any agent with higher priority than t', we have that $|\nu(t''+1) \cap A_{t''+1}| = |\mu(t''+1) \cap A_{t''+1}|$ so $\nu(t''+1) \succsim_{t''+1} \mu(t''+1)$. Moreover, since $\mu \in \mathcal{M}^t(A) \subseteq \mathcal{M}^{t''+1}(A)$ and $\nu(t''+1) \succsim_{t''+1} \mu(t''+1)$, we conclude that $\nu \in \mathcal{M}^{t''+1}(A)$ as well. By induction, this completes the proof that $\nu \in \mathcal{M}^{t'-1}(A)$.

To complete the proof of the Lemma, note that $\nu \in \mathcal{M}^{t'-1}(A)$ and the assumption that C increases the welfare of t', jointly imply $\mu \notin \mathcal{M}^{t'}(A)$. Given that $t' \leq t$, we thus obtain a contradiction to $\mu \in \mathcal{M}^t(A)$. □

In the remainder of this proof, we establish the sufficiency part of Theorem 2 by showing that if $\mu \in \mathcal{M}^{t-1}(A) \setminus \mathcal{M}^t(A)$, then there is a CIC of μ for t at A. Fix some $\mu \in \mathcal{M}^{t-1}(A) \setminus \mathcal{M}^t(A)$. By definition of $\mathcal{M}^t(A)$, $\mu \in \mathcal{M}^{t-1}(A) \setminus \mathcal{M}^t(A)$ implies that there is some $\nu \in \mathcal{M}^{t-1}(A)$

such that $\nu(t) \succ_t \mu(t)$. We construct a CIC of μ for t at A. We start with $i^1 = t$. Since $\nu(t) \succ_t \mu(t)$ and $\nu(t) \cup \mu(t) \subseteq A_t \cup \Omega_t$, we must have $|\nu(t) \cap A_t| > |\mu(t) \cap A_t|$. Let $o^1 \in [\nu(t) \cap A_t] \setminus \mu(t)$.

Suppose now that we have grown a sequence $(i^1, o^1, \ldots, i^M, o^M)$ of M (not necessarily distinct) agents and distinct objects in a way that satisfies the following four conditions:

$$\text{For each } m \geq 1, o^m \in \nu(i^m) \setminus \mu(i^m), \text{ and, for each } m \geq 2, o^{m-1} \in \mu(i^m) \setminus \nu(i^m). \tag{C1}$$

$$\text{For each } m \geq 1, o^m \in \Omega_{i^m} \cup A_{i^m}. \tag{C2}$$

$$\begin{gathered}\text{For each pair } m \text{ such that } i^m < t, \text{ either} \\ \{o^{m-1}, o^m\} \subseteq A_{i^m} \text{ or } \{o^{m-1}, o^m\} \subseteq \Omega_{i^m} \setminus A_{i^m}.\end{gathered} \tag{C3}$$

$$\text{For each } m \geq 2 \text{ such that } i^m = t, \{o^{m-1}, o^m\} \subseteq A_t. \tag{C4}$$

Note first that we can use the sequence $(i^1, o^1, \ldots, i^M, o^M)$ to construct a CIC of μ for t at A if $o^M \in \mu(t) \setminus A_t$: letting m' denote the largest integer strictly smaller than M such that $i^{m'} = t$,[33] $(i^{m'}, o^{m'}, \ldots, i^M, o^M)$ is a CIC of μ for $i^1 = t$ at A.

Next, note that (C1) to (C4) are satisfied for $M = 1$ by definition of o^1. In the remainder of the proof, we show that, whenever $o^M \notin \mu(t) \setminus A_t$, then we can find an agent-object pair (i^{M+1}, o^{M+1}) so that $(i^1, o^1, \ldots, i^{M+1}, o^{M+1})$ satisfies (C1) to (C4). Since the sets of agents and objects are finite, we must thus eventually encounter an M such that a part of $(i^1, o^1, \ldots, i^M, o^M)$ forms a CIC of μ for $i^1 = t$ at A.

Since both ν and μ are matchings, $\sum_{i \in I} |\nu(i)| = \sum_{i \in I} |\mu(i)| = \sum_{i \in I} |\Omega_i|$. Hence, $o^M \notin \mu(i^M)$ implies that there exists an agent $i^{M+1} \in I \setminus \{i^M\}$ such that $o^M \in \mu(i^{M+1})$. Since $o^M \in \nu(i^M)$, we have $o^M \in \mu(i^{M+1}) \setminus \nu(i^{M+1})$. We assume that if $i^{M+1} = t$, then $o^M \in A_t$. We distinguish three cases according to the identity of i^{M+1}. In each of these cases, we identify an object o^{M+1} such that the sequence $(i^1, o^1, \ldots i^{M+1}, o^{M+1})$ satisfies (C1) to (C4).

Case 1: $i^{M+1} < t$ Assume first that $o^M \in A_{i^{M+1}}$. Since objects in $\{o^1, \ldots, o^M\}$ are all distinct, (C3) implies

$$\begin{gathered}|(\nu(i^{M+1}) \setminus \mu(i^{M+1})) \cap \{o^1, \ldots, o^M\} \cap A_{i^{M+1}}| \\ < \\ |(\mu(i^{M+1}) \setminus \nu(i^{M+1})) \cap \{o^1, \ldots, o^M\} \cap A_{i^{M+1}}|.\end{gathered}$$

Thus, combining the last two observations, there exists $o^{M+1} \in [\nu(i^{M+1}) \setminus (\mu(i^{M+1}) \cup \{o^1, \ldots, o^M\})] \cap A_{i^{M+1}}$. The argument in case of $o^M \in \Omega_{i^{M+1}} \setminus A_{i^{M+1}}$ is completely analogous.

[33] If $o^M \in \mu(i^M)$, then (C1) implies that $i^M \neq t$.

Case 2: $\boldsymbol{i^{M+1} > t}$ Since ν and μ are both feasible, $|\mu(i^{M+1})\setminus\nu(i^{M+1})| = |\nu(i^{M+1})\setminus\mu(i^{M+1})|$. However, since $o^M \in \mu(i^{M+1}) \setminus \nu(i^{M+1})$ and for each $m \in \{2, \ldots, M-1\}, o^{m-1} \in \mu(i^m) \setminus \nu(i^m)$ as well as $o^m \in \nu(o^m) \setminus \mu(o^m)$, we have that

$$|(\nu(i^{M+1}) \setminus \mu(i^{M+1})) \cap \{o^1, \ldots, o^M\}| < |(\mu(i^{M+1}) \setminus \nu(i^{M+1})) \cap \{o^1, \ldots, o^M\}|.$$

Thus, there is $o^{M+1} \in (\nu(i^{M+1})\setminus\mu(i^{M+1}))\setminus\{o^1, \ldots, o^M\}$. Given that ν is component-wise individually rational, we have that $o^{M+1} \in A_{i^{M+1}} \cup \Omega_{i^{M+1}}$.

Case 3: $\boldsymbol{i^{M+1} = t}$ Recall that $o^1 \in A_t$ and that (C4) implies, for all $m \geq 2$ such that $i^m = t$, $\{o^{m-1}, o^m\} \subseteq A_t$. Hence,

$$|[\nu(t) \setminus \mu(t)] \cap A_t \cap \{o^1, \ldots, o^M\}| = |[\mu(t) \setminus \nu(t)] \cap A_t \cap \{o^1, \ldots, o^M\}|.$$

Since $|\nu(t) \cap A_t| > |\mu(t) \cap A_t|$, $(\nu(t) \cap A_t) \setminus (\mu(t) \cup \{o^1, \ldots, o^M\}) \neq \emptyset$ and we let o^{M+1} be any element of that set.

$\square$

B Proofs from Section 4

We start with some general observations and auxiliary results before establishing Lemma 3 and Lemma 4. Let i, A, and $o \in O \setminus A_i$. Let $\hat{A}_i = A_i \cup \{o\}$ and $\hat{A} = (\hat{A}_i, A_{-i})$. Assume that $K^i(\hat{A}) \neq K^i(A)$. We show that, for each $\mu \in \mathcal{M}(A)$, there is a cycle C such that

1. i gets o—that is, $o \in (\mu + C)(i) \setminus \mu(i)$, and
2. $\mu + C$ is a CIRP outcome at $\hat{A}$—that is, $\mu + C \in \mathcal{M}^n(\hat{A})$.

Below, we show that the existence of a cycle with the aforementioned properties implies both, Lemma 3 and Lemma 4.

Let j be the highest priority agent for whom $K^j(\hat{A}) \neq K^j(A)$. Note that $K^i(A) \neq K^i(\hat{A})$ implies $j \leq i$.

We first prove that after the j^{th} step of CIRP, i receives o at all remaining matchings.

Claim 1. *For each $\mu \in \mathcal{M}^j(\hat{A})$, $o \in \mu(i)$.*

Proof. We first show via induction on k that, for all $k \leq j$, $\mathcal{M}^{k-1}(A) \subseteq \mathcal{M}^{k-1}(\hat{A})$ and $o \in \mu(i)$ for each $\mu \in \mathcal{M}^{k-1}(\hat{A}) \setminus \mathcal{M}^{k-1}(A)$. For $k = 1$, both statements follow from $\hat{A}_i = A_i \cup \{o\}$

and $\hat{A}_l = A_l$ for all $l \neq i$. So consider some $k \in \{2, \ldots, j\}$ and assume that both statements have been shown for all $k' < k$. We have that

$$\begin{aligned}
\mathcal{M}^{k-1}(A) &= \{\mu \in \mathcal{M}^{k-2}(A) : |\mu(k-1) \cap A_{k-1}| = K^{k-1}(A)\} \\
&\subseteq \{\mu \in \mathcal{M}^{k-2}(\hat{A}) : |\mu(k-1) \cap A_{k-1}| = K^{k-1}(A)\} \\
&= \{\mu \in \mathcal{M}^{k-2}(\hat{A}) : |\mu(k-1) \cap \hat{A}_{k-1}| = K^{k-1}(\hat{A})\} \\
&= \mathcal{M}^{k-1}(\hat{A})
\end{aligned}$$

Here, the subset relation follows from the inductive assumption that $\mathcal{M}^{k-2}(A) \subseteq \mathcal{M}^{k-2}(\hat{A})$ and the second equality follows from the definitions of j and $\hat{A}$ since $k-1 < j \leq i$. In order to show the second part of the statement for k, fix $\mu \in \mathcal{M}^{k-1}(\hat{A})$ such that $o \notin \mu(i)$. By construction of $\mathcal{M}^{k-1}(\hat{A})$, we have $\mu \in \mathcal{M}^{k-2}(\hat{A})$. Since $o \notin \mu(i)$, the inductive assumption for $k-1$ implies $\mu \in \mathcal{M}^{k-2}(A)$. Since $k-1 < j \leq i$, we have that $|\mu(k-1) \cap A_{k-1}| = |\mu(k-1) \cap \hat{A}_{k-1}|$ and $K^{k-1}(A) = K^{k-1}(\hat{A})$. Combining the last two statements, we obtain $\mu \in \mathcal{M}^{k-1}(A)$.

We now use the just established statements for $k = j$ to complete the proof. Since $\mathcal{M}^{j-1}(A) \subseteq \mathcal{M}^{j-1}(\hat{A})$, $K^j(\hat{A}) \geq K^j(A)$ and the definition of j implies $K^j(\hat{A}) > K^j(A)$. Thus, $\mathcal{M}^j(\hat{A}) \subseteq \mathcal{M}^{j-1}(\hat{A}) \setminus \mathcal{M}^{j-1}(A)$. Hence, $o \in \mu(i)$ for all $\mu \in \mathcal{M}^j(\hat{A})$. □

Let $\mu \in \mathcal{M}^n(A)$. By Theorem 2, since $\mu \notin \mathcal{M}^j(\hat{A})$, there exists a CIC of μ for j at $\hat{A}$. Let $\mathcal{C}^j$ be the set of all such CICs.

Claim 2. *In each $C \in \mathcal{C}^j$, i gets o.*

Proof. If not, then C is a CIC of μ for j at A, contradicting $\mu \in \mathcal{M}^j(A)$. □

Now, for each $l > j$, let $\mathcal{C}^l$ consist of the cycles of μ in $\mathcal{C}^{l-1}$ that are most beneficial for l. That is, let

$$\mathcal{C}^l = \{C \in \mathcal{C}^{l-1} : \mu + C \succsim_l \mu + C' \text{ for all } C' \in \mathcal{C}^{l-1}\}$$

The following claim is the heart of the argument for both Lemmas 3 and 4.

Claim 3. *For each $l \geq j$,*

$$\{\mu + C : C \in \mathcal{C}^l\} \subseteq \mathcal{M}^l(\hat{A}).$$

Proof. Assume to the contrary that Claim 3 is false and let L be the smallest integer greater than or equal to j such that the statement fails. Formally, suppose that L is such that there is $C \in \mathcal{C}^L$ such that $\mu + C \notin \mathcal{M}^L(\hat{A})$ and for each $\tilde{C} \in \mathcal{C}^{L-1}$, $\mu + \tilde{C} \in \mathcal{M}^{L-1}(\hat{A})$. By Theorem 2, there exists a CIC C' of $\mu + C$ for L at $\hat{A}$. Let $\nu = (\mu + C) + C'$.

By Lemma 2 there is a decomposition of $\nu - \mu$. For the remainder of this proof, we fix this decomposition as $\{C_1, C_2, \ldots C_K\}$. It is without loss of generality to suppose that $o \notin O(C_2) \cup \cdots \cup O(C_K)$.

Next, we argue that $K \geq 2$. If C and C' both increase the welfare of L, the statement is immediate. If C does not affect L and $K = 1$, we have $\mu + C_1 = \mu + C + C'$. Since C does not affect L and C' increases the welfare of L, C_1 must increase the welfare of L. Since C' cannot affect any agent $l < L$ (as C' is a CIC for L), each such agent is indifferent between $\mu + C_1$ and $\mu + C + C'$. Hence, we have $C_1 \in \mathcal{C}^L$ and obtain a contradiction to our assumption that $C \in \mathcal{C}^L$ since $\mu + C_1 \succ_L \mu + C$. Finally, if C decreases the welfare of L and $K = 1$, we have $\mu + C_1 = \mu + C + C'$ and C_1 does not affect the welfare of L. As in the previous case, all agents $l < L$ are indifferent between $\mu + C_1$ and $\mu + C + C'$, so that we obtain $C_1 \in \mathcal{C}^L$. Since $\mu + C_1 \succ_L \mu + C$, we obtain another contradiction to $C \in \mathcal{C}^L$.

We now argue inductively that, for any t, there exists a sequence of agents $(k^0, \ldots, k^t)$ and indices $(m^0, \ldots, m^t)$ such that

1. $k^t < \cdots < k^0 < L$
2. If $m^t = 1$, then either
   (a) $C_{m^t}(= C_1)$ increases the welfare of k^t and C does not affect k^t, or
   (b) $C_{m^t}(= C_1)$ does not affect k^t, C decreases the welfare of k^t, and there is no m such that C_m increases the welfare of k^t.
3. If $m^t \neq 1$, then C_{m^t} increases the welfare of k^t.

Since such a sequence exists for any finite t, we eventually obtain a contradiction to our assumption that $|I| = n$.

For the induction base, we distinguish three cases according to how C affects the welfare of L.

Case 1: C increases the welfare of L. Since C and C' increase the welfare of L, we assume, without loss of generality, that C_2 increases L's welfare. Since $o \notin O(C_2)$ and $\mu \in \mathcal{M}^n(A)$, there must be some $k^0 < L$ such that C_2 decreases k^0's welfare; otherwise C_2 would be a CIC of μ for L at A, contradicting $\mu \in \mathcal{M}^n(A) \subseteq \mathcal{M}^L(A)$. If there is no cycle that increases the welfare of k^0, then C decreases the welfare of k^0 and we let $m^0 = 1$ to obtain a pair (k^0, m^0) that satisfies the three conditions above. If there exists m^0 such that C_{m^0} increases the welfare of k^0, then C does not affect the welfare of k^0 and (k^0, m^0) is a pair with the desired properties.[34]

[34] Since $k^0 < L$, C' does not affect k^0. Since C_2 decreases k^0's welfare and C_{m^0} increases it, the net effect of C_2 and C_{m_0} is to leave k^0 unaffected. By definition of the decomposition, this net effect is equal to the net effect of C and C'. Thus, since C' does not affect k^0, we infer that C does not affect k^0 either. Analogous comments apply to other cases in which we encounter an agent with higher priority than L whose welfare is decreased by one of the cycles in the decomposition.

Case 2: C does not affect L. Since C' is a CIC of $\mu + C$ for L, there exists m such that C_m increases the welfare of L.

Subcase 2.1: $m = 1$. There is $k^0 < L$ who strictly prefers C over C_1 since we otherwise contradict $C \in \mathcal{C}^L$. There exists a unique $m^0 \neq 1$ such that C_{m^0} increases the welfare of k^0 and (k^0, m^0) is a pair with the desired properties.[35]

Subcase 2.2: $m \neq 1$. There is $k^0 < L$ whose welfare is decreased by C_m since, given that $o \notin O(C_m)$, C_m would be a CIC of μ for L at A otherwise. If there is no cycle that increases the welfare of k^0, then C decreases the welfare of k^0 and we let $m^0 = 1$ to obtain a pair (k^0, m^0) that satisfies the three conditions above.[36] If there exists m^0 such that C_{m^0} increases the welfare of k^0, then C does not affect the welfare of k^0 and (k^0, m^0) is a pair with the desired properties.

Case 3: C decreases the welfare of L. Since C' is a CIC of $\mu + C$ for L, L is indifferent between μ and $(\mu + C) + C'$. Since C decreases the welfare of L, there either exists m such that C_m increases the welfare of L, or there is no m such that C_m affects the welfare of L.

If there is $m \neq 1$ such that C_m increases the welfare of L, we proceed as in Subcase 2.2. If there is no such m, we proceed as in Subcase 2.1, no matter whether C_1 increases the welfare of L or leaves her unaffected, we have $\mu + C_1 \succ_L \mu + C$.

Having established the base case, suppose that we have constructed a pair of sequences $(k^0, \ldots, k^t)$ and $(m^0, \ldots, m^t)$ with the three properties introduced above. We now identify k^{t+1} and m^{t+1} to extend the two sequences in such a way that preserves the desired properties.

Case 1: C_{m^t} increases the welfare of k^t and $m^t \neq 1$. Since $o \notin O(C_{m^t})$, there must be some $k^{t+1} < k^t$ such that C_{m^t} decreases k^{t+1}'s welfare—otherwise, we obtain a contradiction to $\mu \in \mathcal{M}^n(A)$ since C_{m^t} would be a CIC of μ for k^t at A. If there is no cycle that increases the welfare of k^{t+1}, then C decreases the welfare of k^{t+1} and we let $m^{t+1} = 1$. Otherwise, there exists m^{t+1} such that $C_{m^{t+1}}$ increases the welfare of k^{t+1}, while C does not affect k^{t+1}. In either case, the pair (k^{t+1}, m^{t+1}) extends our sequence.

[35] Since $k^0 < L$, C' does not affect k^0. If C increases the welfare of k^0 but C_1 does not, then $\nu(k^0) \succ_{k^0} \mu(k^0)$ and there must therefore exist a unique m^0 that increases the welfare of k^0. If C does not affect k^0, then k^0 prefers C over C_1 if and only if C_1 decreases the welfare of k^0. Since neither C nor C' affect k^0, we must have $\nu(k^0) \sim_{k^0} \mu(k^0)$ and, given that C_1 decreases the welfare of k^0, there must exist a unique $m^0 \neq 1$ such that C_{m^0} increases the welfare of k^0. Analogous comments apply to other cases in which we encounter an agent with higher priority than L who strictly prefers C over some cycle in the decomposition of $\nu - \mu$.

[36] Since $k^0 < L$, C' does not affect k^0. Since C decreases the welfare of k^0 and C' does not affect it, there cannot be a cycle different from C_m that affects k^0. Analogous comments apply to other cases in which we encounter an agent with higher priority than L whose welfare is once decreased but never increased by a cycle in the decomposition of $\nu - \mu$.

Case 2: C_{m^t} increases the welfare of k^t and $m^t = 1$. By Property (2a), C does not affect the welfare of k^t. Since $k^t < L$ and $C \in \mathcal{C}^L$, there is $k^{t+1} < k^t$ who strictly prefers C over C_1. But then, there is some $m^{t+1} \neq 1$ such that $C_{m^{t+1}}$ increases the welfare of k^{t+1}. The pair (k^{t+1}, m^{t+1}) extends our sequence.

Case 3: C_{m^t} does not affect the welfare of k^t. By Property (2b), $m^t = 1$ and C decreases the welfare of k^t. Hence, $\mu + C_1 \succ_{k^t} \mu + C$ and, given that $C \in \mathcal{C}^L \subseteq \mathcal{C}^{k^t}$ (as $k^t < L$), there is $k^{t+1} < k^t$ who strictly prefers C over C_1. Since C_1 decreases k^{t+1}'s welfare, there is $m^{t+1} \neq 1$ such that $C_{m^{t+1}}$ increases the welfare of k^{t+1} and we extend our sequence by the pair (k^{t+1}, m^{t+1}).

□

Proof of Lemma 3. If $K^i(\hat{A}) = K^i(A)$, there is nothing left to show. So assume that $K^i(\hat{A}) \neq K^i(A)$. By Claim 3, we immediately obtain $K^i(\hat{A}) \leq |\mu(i) \cap A_i| + 1 = K^i(A) + 1$. Let $\nu \in \mathcal{M}^n(\hat{A})$. By Claim 2, if $K^i(\hat{A}) \neq K^i(A)$, then we have $o \in \nu(i)$. Thus, since $o \in \hat{A}_i \setminus A_i$, $|\nu(i) \cap A_i| = |\nu(i) \cap \hat{A}_i| - 1 = K^i(\hat{A}) - 1 \leq K^i(A)$. □

Proof of Lemma 4. We now reverse the roles of A_i and $\hat{A}_i$: let $\hat{A}_i$ be the true set of desirable objects for agent i and $A_i = \hat{A}_i \setminus \{o\}$ be a possible misreport for agent i. Again, if $K^i(\hat{A}) = K^i(A)$, then there is nothing left to show. So assume that $K^i(\hat{A}) \neq K^i(A)$. Fix a matching $\mu \in \mathcal{M}^n(A)$. By Claim 3, there exists a cycle C such that $\mu + C \in \mathcal{M}^n(\hat{A})$. By Claim 2, i must obtain o in C. Since $o \in \hat{A}_i$, C cannot decrease the welfare of i when i's true set of desirable objects is $\hat{A}_i$. But then, $K^i(\hat{A}) \neq K^i(A)$ implies $K^i(\hat{A}) > K^i(A)$ and the deviation from $\hat{A}_i$ to $A_i = \hat{A}_i \setminus \{o\}$ is not profitable for i. □

Before concluding with the proof of Theorem 3, we strengthen Lemma 3 to consider addition of not only a single object, but a set of objects to an agent's desirable set.

Lemma 7. *Let $A \in \mathcal{A}$ and $B \subseteq O \setminus A_i$. Denote by A^B the profile $(A_i \cup B, A_{-i})$. For each $\mu^B \in \mathcal{M}^i(A^B)$, we have $|\mu^B(i) \cap A_i| \leq K^i(A)$.*

Proof. We proceed by induction on the cardinality of B. Lemma 3 establishes the base case when $|B| = 1$. Suppose that for some $L \geq 2$ the statement of the Lemma has been established for all pairs A and B such that $|B| < L$. Let $A \in \mathcal{A}$ and $B \subseteq O \setminus A_i$ be such that $|B| = L$ and let $B = \{o^1, \ldots, o^L\}$. Let $\mu^B \in \mathcal{M}^i(A^B)$. For the sake of contradiction, suppose $|\mu^B(i) \cap A_i| > K^i(A)$. Let l be the smallest integer such that $o^l \in \mu^B(i)$, $A_i^{\geq l} = A_i \cup \{o^l, \ldots, o^L\}$, and $A^{\geq l} = (A_i^{\geq l}, A_{-i})$. By definition of CIRP, $\mu^B \in \mathcal{M}^i(A^{\geq l})$. Next, let $A_i^l = A_i \cup \{o^l\}$, $A^l = (A_i^l, A_{-i})$, and $\mu^l \in \mathcal{M}^i(A^l)$. By the inductive assumption, $|\mu^l(i) \cap A_i| \leq K^i(A)$. Hence,

$|\mu^l(i) \cap (A_i \cup \{o^l\})| \leq K^i(A) + 1$. Since we assumed that $|\mu^B(i) \cap A_i| > K^i(A)$ and $o^l \in \mu^B(i)$, we have that $|\mu^B(i) \cap (A_i \cup \{o^l\})| > |\mu^l(i) \cap (A_i \cup \{o^l\})| = K^i(A^l)$. Hence, we obtain a contradiction to the inductive assumption for the pair A_i^l and $B^l = \{o^{l+1}, \ldots, o^L\}$ since $|B^l| \leq L - 1$. □

Proof of Theorem 3. Let A be a profile of desirable sets and, for some $i \in I$, let $\hat{A}_i$ be an alternative report for i.

Suppose that $\hat{A}_i \setminus A_i \neq \emptyset$. Let $\tilde{A}_i = A_i \cap \hat{A}_i$ and $B = \hat{A}_i \setminus A_i$. Moreover, let $\mu^B \in \mathcal{M}(\tilde{A}_i \cup B, A_{-i}) = \mathcal{M}(\hat{A})$ and $\tilde{\mu} \in \mathcal{M}(\tilde{A}_i, A_{-i})$. Applying Lemma 7, $|\mu^B(i) \cap A_i| = |\mu^B(i) \cap \tilde{A}_i| \leq K^i(\tilde{A}) = |\tilde{\mu}(i) \cap A_i|$. Thus, $\tilde{A}_i \subseteq A_i$ is at least as profitable a manipulation for i as $\hat{A}_i$. Hence, we can proceed with our analysis with $\tilde{A}_i$ in place of $\hat{A}_i$.

Now suppose that $\hat{A}_i \subsetneq A_i$. Let $\{o^1, \ldots, o^T\} = A_i \setminus \hat{A}_i$ for some $T \geq 1$. For any $t \leq T$, let $\hat{A}_i^t = \hat{A}_i \cup \{o^1, \ldots, o^t\}$ and $A^t = (A_i^t, A_{-i})$. By Lemma 4 (taking A_i^1 to be the true and $\hat{A}_i$ to be the false set of desirable objects), we obtain $K^i(A^1) \geq K^i(\hat{A})$. Proceeding inductively, T applications of Lemma 4 yield that $K^i(A^T) \geq K^i(\hat{A})$. Hence, by definition of $o^1, \ldots, o^T$ we obtain that $K^i(A) \geq K^i(\hat{A})$. □

C Proof of Theorem 4

To prove part 1 of the Theorem we show by induction on t that the maximum flows of (V, E, q) correspond to $\mathcal{M}^t(A)$ after the t^{th} iteration of the for-loop on line 2 of Algorithm 1.

As a base case, we show that for (V, E, q), where q is the initial capacity vector defined above, f is a maximum flow if and only if it corresponds to some matching $\mu \in \mathcal{M}^0(A)$. Take any matching $\mu \in \mathcal{M}^0(A)$. The definition of a matching and our restriction to trichotomous preferences imply $|\mu(i)| = |\Omega_i|$ and $\mu(i) \subseteq \Omega_i \cup A_i$. Define a flow f by setting, for all $i \in I$, $f(S, i) = |\Omega_i|$, $f(i, i^A) = |\mu(i) \cap A_i|$, $f(i, i^U) = |\mu(i) \cap (\Omega_i \setminus A_i)|$, $f(i^A, o) = 1$ for all $o \in \mu(i) \cap A_i$, $f(i^U, o) = 1$ for all $o \in \mu(i) \cap (\Omega_i \setminus A_i)$, and, for all $o \in \cup_{i \in I} \mu(i)$, $f(o, T) = 1$. Clearly, f is a maximum flow in (V, E, q). Conversely, take a maximum flow f in (V, E, q) and let μ be the matching that corresponds to f. Since all maximum flows have value $\sum_i |\Omega_i|$, we have $|\mu(i)| = |\Omega_i|$ for all $i \in I$. Since (V, E, q) contains no paths between i and an object in $O \setminus (\Omega_i \cup A_i)$, we have $\mu(i) \subseteq \Omega_i \cup A_i$. Hence, $\mu \in \mathcal{M}^0(A)$.

As an induction hypothesis, suppose that at the start of the t^{th} iteration of the loop, the maximum flows of (V, E, q) correspond to $\mathcal{M}^{t-1}(A)$. Let $\hat{q}$ be the capacity at the end of the t^{th} iteration of the for-loop, f be a maximum flow in $(V, E, \hat{q})$, and μ be the matching that corresponds to f. We show first that $\mu \in \mathcal{M}^t(A)$. By construction, $|\mu(t) \cap A_t| = |\Omega_t| - \hat{q}(i, i^U)$. Assume that there is a matching $\mu' \in \mathcal{M}^t(A)$ such that $|\mu'(t) \cap A_t| > |\mu(t) \cap A_t|$. Since the set of maximum flows of (V,E,q) corresponds to $\mathcal{M}^{t-1}(A)$ and since $\mu' \in \mathcal{M}^t(A)$, just as argued in the base case, μ' induces a maximum flow f' in (V,E,q). Given that $|\mu'(t) \cap A_t| > |\mu(t) \cap A_t|$, f' is a maximum flow in $(V,E,\tilde{q})$, where $\tilde{q}(t,t^U) = \hat{q}(t,t^U) - 1$ and $\tilde{q}(e) = \hat{q}(e)$ for all $e \in E \setminus \{(t,t^U)\}$. But then, the t^{th} iteration of the loop would have concluded with the capacity of (i,i^U) no higher than $\hat{q}(t,t^U) - 1$. This contradiction shows that $\mu \in \mathcal{M}^t(A)$. Conversely, take any matching $\hat{\mu} \in \mathcal{M}^t(A)$ and let $\hat{f}$ be the induced flow, defined as in the base case. By our previous arguments, we have $|\hat{\mu}(t) \cap A_t| = |\mu(t) \cap A_t|$. Since $|\hat{\mu}(t)| = |\mu(t)| = |\Omega_t|$, we obtain that $|\hat{\mu}(t) \setminus A_t| = |\mu(t) \setminus A_t| = \hat{q}(t,t^U)$. Given that $\mu \in \mathcal{M}^t \subseteq \mathcal{M}^{t-1}$ and that $\hat{q}(e) = q(e)$ for all $e \in E \setminus \{(t,t^U)\}$, the induction hypothesis implies that $\hat{f}$ is a maximum flow in $(V,E,\hat{q})$.

We now turn to part 2 of the Theorem. Let $m = |O|$ and $n = |I|$. Then there are at most m edges out of i^A for each $i \in I$, so the Ford-Fulkerson algorithm computes MaxFlow in $O(mn^2)$ time. We invoke MaxFlow while minimizing the flow along (i,i^U), so it suffices to call it $O(log(k))$ times for each agent, where $k = \max_{i \in I} |\Omega_i|$. Thus, the complexity of IRP is $O(mn^3 log(k))$. $\square$

D Proof of Proposition 2

We first specialize the IRP algorithm for all-or-nothing preferences. Given a profile of desirable sets A and a profile of targets k, the algorithm works as follows:

> **Step 0**: Let $\mathcal{M}^0(A,k)$ be the set of all individually rational matchings at the all-or-nothing preferences with respect to (A,k).
>
> **Step $t \in \{1,\dots,n\}$**: Let
>
> $$K^t(A,k) = \begin{cases} k_i & \text{if there is } \mu \in \mathcal{M}^{t-1}(A,k) \text{ such that } |\mu(i) \cap A_i| = k_i \\ |\Omega_i \cap A_i| & \text{otherwise} \end{cases}$$
>
> and
>
> $$\mathcal{M}^t(A,k) = \left\{ \mu \in \mathcal{M}^{t-1}(A,k) : \begin{array}{ll} |\mu(t) \cap A_t| = K^t(A,k) & \text{if } K^t(A,k) = k_t \text{ and} \\ \mu(t) = \Omega_t & \text{otherwise} \end{array} \right\}.$$

Now, fix a profile of desirable sets, A, and a profile of targets, k. Let $i \in I$. By definition of all-or-nothing preferences, it is impossible for i to gain by reporting a distinct target $\hat{k}_i \neq k_i$. Thus, we show that i does not benefit by reporting an alternative desirable set $\hat{A}_i \neq A_i$.

If $\hat{A}_i$ is a *profitable* manipulation for i, then $|\Omega_i \cap A_i| < k_i$, $\mu(i) = \Omega_i$ for each $\mu \in \mathcal{M}^i(A, k)$, and there has to exist a $\hat{\mu} \in \mathcal{M}^i(\hat{A}_i, A)$ such that $|\hat{\mu}(i) \cap A_i| = k_i$ and $\hat{\mu}(i) \subseteq \Omega_i \cup A_i$.

Suppose there is $t < i$ such that $K^t(A, k) \neq K^t((\hat{A}_i, A_{-i}), k)$. Let t be the first of such agents. Since for each $t' < t$, $K^{t'}(A, k) = K^{t'}((\hat{A}_i, A_{-i}), k)$, we draw the following two conclusions:

1. Since $\mu(i) = \Omega_i$ and $\mu \in \mathcal{M}^{t-1}(A, k)$, the definition of t implies $\mu \in \mathcal{M}^{t-1}((\hat{A}_i, A_{-i}), k)$. This, in turn, implies that

$$K^t((\hat{A}_i, A_{-i}), k) \geq |\mu(t) \cap A_t| = K^t(A, k).$$

2. Since $|\hat{\mu}(i) \cap A_i| = k_i$, the definition of t implies $\hat{\mu} \in \mathcal{M}^{t-1}(A, k)$. This, in turn, implies that

$$K^t(A, k) \geq |\hat{\mu}(t) \cap A_t| = K^t((\hat{A}_i, A_{-i}), k).$$

Together, we have a contradiction of the supposition that $K^t(A, k) \neq K^t((\hat{A}_i, A_{-i}), k)$. Thus, for each $t < i$, $K^t(A, k) = K^t(\hat{A}_i, A_{-i})$, implying that $\hat{\mu} \in \mathcal{M}^{i-1}(A, k)$. So $K^i(A, k) \geq K^i((\hat{A}_i, A_{-i}), k)$, contradicting the assumption that $\hat{A}_i$ is a profitable manipulation for i. $\square$