Efficient Assignment with Interdependent Values*

Yeon-Koo Che, Jinwoo Kim, Fuhito Kojima†

April 23, 2012

Abstract: We study the "house allocation" problem in which n agents are assigned n objects, one for each agent, when the agents have interdependent values. We show that there exists no mechanism that is Pareto efficient and ex post incentive compatible, and the only mechanism that is group ex post incentive compatible is constant across states. By contrast, we demonstrate that a Pareto efficient and Bayesian incentive compatible mechanism exists in the 2 agent house-allocation problem, given sufficient congruence of preferences and the standard single crossing property.

*We thank So Yoon Ahn, Yuchen Liu, Kelly Rader, Keeyoung Rhee, and especially Fanqi Shi for excellent research assistance. We also thank Oguz Afacan, Eduardo Azevedo, Dirk Bergemann, Wing-Sum Cheung, Jon Eguia, Daniel Fragiadakis, Tadashi Hashimoto, Navin Kartik, Konrad Mierendorff, Benny Moldovanu, Bruno Strulovici, and seminar participants at Seoul National University for comments. Yeon-Koo Che and Jinwoo Kim acknowledge the financial support from the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology through its World Class University Grant (R32-2008-000-10056-0). This work was completed while Jinwoo Kim and Fuhito Kojima were visiting Yonsei and Columbia University, whose support and hospitality are gratefully acknowledged.

†Che: Department of Economics, Columbia University, and YERI, Yonsei University (email: yeonkooche@columbia.edu); Kim: Department of Economics, Seoul National University (email: jikim72@gmail.com); Kojima: Department of Economics, Stanford University (fuhitokojima1979@gmail.com).

1 Introduction

Many real life allocation problems involve assigning indivisible objects to individuals without monetary transfer. Examples include university housing allocation, office assignment, and student placement in public schools. A typical goal in such a problem is to assign the objects efficiently while eliciting true preferences of the participants. The literature on matching and market design has made considerable advances under the assumption that agents have private values, namely that participants know the values of the objects being assigned (at least in expectation). Given the private-value assumption, studies in this literature have identified a number of mechanisms that implement a Pareto efficient allocation in a **strategy-proof** fashion, making it a dominant strategy for each participant to reveal his preferences truthfully.[1] These and related studies have helped the redesign of school choice programs in cities such as Boston and New York City.[2]

In many resource allocation problems, however, participants do not possess sufficient information about the objects being allocated. School choice is a case in point. A major challenge facing students and their parents in school choice is that they do not know enough about schools to form clear preferences about them. They consult school websites, information booths, fairs and campus tours. But they also seek advice from others through word-of-mouth, online social networks, and guidebooks, and often get swayed by the anecdotes and personal experiences they are told.[3]

[1]Examples of mechanisms with these features include serial dictatorships (Svensson, 1999; Abdulkadiroğlu and Sönmez, 1998), top trading cycles mechanisms (Abdulkadiroğlu and Sönmez, 2003), hierarchical exchanges (Papai, 2000), and trading cycles mechanisms (Pycia and Ünver, 2009).

[2]See Abdulkadiroğlu, Pathak and Roth (2005), and Abdulkadiroğlu et al. (2005) who helped design student placement mechanisms in New York City and Boston.

[3]For instance, high school applicants "constantly talk about which colleges each high school sends its graduates to, where there might be more interesting students, how long the subway ride would be." (see "Even an Expert's Resolve Is Tested by the City's High School Admissions Process," *New York Times*, December 8, 2008). The importance of the information (or lack thereof) about schools also appears to be behind the immense popularity of websites such as GreatSchools.org, RateMyProfessors.com, and Insideschools.org. The first two websites enjoy more than 800,000 and 13,000,000 ratings and reviews on schools and college professors by students and their parents, respectively. There are several influential guide books, such as *New York City's Best Public Schools* series, written by Clara Hemphill, which is "regarded as the bible for navigating school choices," according to the aforementioned article.

The scenario described here departs starkly from the private-value setting portrayed by most existing studies in matching and market design.[4] Instead, dispersed information and the relevance of local information and others' personal experiences make parents' preferences interdependent.[5] That is, a parent's information affects the preferences of other parents. Interdependence of preferences is also present in other allocation problems, such as student housing assignments, course allocation, after-school program assignment, and others. How should one design allocation mechanisms in such an environment? How does preference interdependence affect the performance of allocation mechanisms?

In private-values settings, strategy-proofness has been regarded as a desirable property since it ensures that participants are not harmed by reporting their preferences truthfully, irrespective of their beliefs about other players. Unfortunately, preference interdependence makes strategy-proofness virtually impossible to attain. A natural adaptation of the strategy-proofness concept is **ex post incentive compatibility**, which requires that truth-telling form mutual best responses for every signal profile. Much like strategy-proofness, ex-post incentive compatibility makes it safe for participants to report signals truthfully, by making truth-telling a best response irrespective of the types of other agents.[6]

Our main finding is that such robustness comes with a heavy price. We show that there exists no mechanism that is Pareto efficient and ex post incentive compatible whenever non-trivial preference interdependence exists (and the preference domain is sufficiently

[4]In private-values models, parents have clear preferences about their choices but are concerned about how to "play" the application game. For many parents, a more difficult problem is to determine what school is good for their child. To see how differently a parent in this latter scenario would behave relative to the one in the former scenario, suppose in a Boston mechanism, a parent receives a word-of-mouth information suggesting that many other parents view a given school as desirable. According to the viewpoint from the existing theory (first scenario), the parent will more likely respond to that information by avoiding ranking that school at the top of her list. But the parent in the latter scenario may more likely rank it at the top, realizing that the school is actually good.

[5]The term "interdependence" refers to informational externalities, namely, that one's value of an object depends on the private information held by others. Importantly, it does not include allocative externalities — namely, that one's preference depends on the other agents' assignments, as would be the case with peer effects.

[6]Given its appeal, the concept of ex post incentive compatibility is used extensively in mechanism design. See Bergemann and Välimäki (2002), Cremer and McLean (1985), Esö and Maskin (2002), Krishna (2003), and Perry and Reny (2002) for instance.

rich). Further, if we require the mechanism to be ex post "group" incentive compatible—namely that there be no group of agents who can benefit from joint manipulations—, we find that only a trivial allocation that prescribes a constant outcome across states can be implemented. These negative results hold even when the value interdependence is arbitrarily small so the preferences are nearly private, which stands in stark contrast to efficiency obtained in "pure" private value models.

Finally, we show that weakening the ex post requirements can lead to more desirable allocations. More specifically, in a setting with two agents and two objects, a Pareto efficient and Bayesian incentive compatible mechanism exists if the standard single crossing property holds and agents' preferences are sufficiently congruent. Our analysis suggests that it may be important to pay attention to mechanisms that violate ex post incentive compatibility but satisfy Bayesian incentive compatibility in order to achieve societal goals if interdependence of valuations exists. This is in a sharp contrast to private-values setting, in which various studies in recent matching and market design literature have emphasized the importance of strategy-proofness (see Abdulkadiroğlu, Pathak and Roth (2009) for instance).

2 Related Literature

Our findings intersect with several strands of existing research. First, the central theme of our paper agrees with Jehiel and Moldovanu (2001) and Jehiel et al. (2006) who investigate the difficulties associated with interdependent values under the transferable utility setup. Specifically, the former paper establishes generic impossibility of implementing the efficient allocation in Bayesian equilibrium; and the latter proves the generic impossibility of implementing an allocation that varies nontrivially with states in ex post equilibrium. While our results reinforce and complement these papers, there are several important distinctions.

First, the inefficiency result of Jehiel and Moldovanu (2001) (established in Bayesian implementation) may at first glance appear to imply ours (established in ex post implementation, which is a stronger requirement), but the efficiency requirements are different between the two models. Specifically, they employ utilitarian efficiency as the welfare criterion, whereas we focus on Pareto efficiency.[7] The latter is much weaker, and there are often many Pareto efficient allocations. Hence, to show the impossibility of efficiency, one must show that *all* such allocations are unattainable. Second, unlike the public decision problem Jehiel et al. (2006) consider,[8] our triviality result is derived in the private-object setting. In the private-object setting, each agent is indifferent across a number of allocations as long as her own assignment is identical. As shown by Bikhchandani (2006), this fact can be exploited to provide non-trivial mechanisms in the private-object setting (with monetary transfers available). Finally, the results of both Jehiel and Moldovanu (2001) and Jehiel et al. (2006) require that agents have multi-dimensional signals while ours do not. Due to these distinctions, our impossibility results are not implied by these papers, but rather extend their insights to a non-transferable utility environment.

Our model is also related to Chakraborty, Citanna and Ostrovsky (2010) and Chakraborty and Citanna (2011), who study preference interdependence in a matching context. Their setup deals with two-sided matching in which agents on one side are matched with agents on the other side, whereas agents are assigned objects in our setup. This difference entails crucial distinctions both in terms of the problems studied and the main thrust of the analysis. For instance, the primary concern in their paper is stability of matching between the two sides, whereas our primary focus is on the efficiency of allocations.[9]

Finally, the current study is part of a growing research field of matching and market design. Gale and Shapley (1962) formalized the two-sided matching problem, and Roth (1984) stimulated early applications of matching theory to economic problems. In particular, market design for student placement due to Balinski and Sönmez (1999) and Abdulkadiroğlu and Sönmez (2003) has been extensively studied in recent years. As mentioned above, the main difference of the current paper from this line of studies is our attention to interde-

[7] The reason for the difference is the environments that these two papers focus on. Jehiel and Moldovanu (2001) consider a transferable utility environment in which Pareto efficiency boils down to utilitarian efficiency. Utilitarian efficiency is not implied by Pareto efficiency, however, in our non-transferable utility environment.

[8] By contrast, the impossibility of efficiency continues to hold in the private object setting (see Example 14 of Jehiel and Moldovanu (2006)). We thank Benny Moldovanu for informing us of this result.

[9] While Chakraborty and Citanna (2011) also consider efficiency, two-sidedness of matching makes their notion quite distinct from ours. They assume the agents on one side have common preferences of the agents on the other side, so every non-wasteful (full) matching is Pareto efficient.

pendent values. The field is too large to summarize here. Instead, we refer to surveys of the literature by Roth and Sotomayor (1990), Roth (2002), Sönmez and Ünver (2009), and Pathak (2011).

3 Illustrative Example

We illustrate the main insight for our impossibility results in a simple setup in which two agents, 1 and 2, are assigned two objects, a and b, one for each agent. There is no money in this economy. Let $v_o^i(s^1, s^2) > 0$ denote agent $i = 1, 2$'s value of receiving object $o = a, b$, when agent $j = 1, 2$ has signal $s^j \in [0, 1]$. Without loss, consider agent i's net utility gain $u^i(s) := v_a^i(s) - v_b^i(s)$ from receiving a instead of b, when the signal profile is $s = (s^1, s^2)$. The function u^i for each agent $i = 1, 2$ is increasing in both signals and satisfies the single crossing property:

$$\frac{\partial u^i(s)}{\partial s^i} > \frac{\partial u^{-i}(s)}{\partial s^i}, \forall s \in [0,1]^2, \tag{1}$$

that is, one's signal affects his own value more than the other's.

Let $S_{o_1 o_2}$ denote the set of signal profiles such that agent 1 prefers object $o_1 \in \{a, b\}$ and agent 2 prefers object $o_2 \in \{a, b\}$, strictly for at least one agent.[10] A **Pareto efficient** allocation must assign a to 1 and b to 2 when the signal profile is in S_{ab} (because 1 likes a more than b, and 2 likes b more than a in S_{ab}), and likewise must assign a to 2 and b to 1 when the signal profile is in S_{ba}. Assume that both of these sets are nonempty. These sets are depicted as shaded areas, respectively, in Figure 1. (In this figure, agent i's indifference curve depicts the locus of signal profiles that make her indifferent between the two objects; i.e., the set $\{s \in [0, 1]^2 | u^i(s) = 0\}$.) Note that Pareto efficiency does not uniquely determine the assignment when both agents prefer a to b (i.e., when the signal profile is in S_{aa}) or when both agents prefer b to a (i.e., when the signal profile is in S_{bb}), or when both of them are indifferent.

We first show that there exists no mechanism that is Pareto efficient and ex post incentive compatible.[11] To see this, suppose otherwise. Then, by the revelation principle, there

[10] For instance, $S_{ab} := \{s \in [0,1]^2 \mid u^1(s) > 0, u^2(s) \le 0 \text{ or } u^1(s) \ge 0, u^2(s) < 0\}$.

[11] As will be seen, the notion of ex post incentive compatibility must be defined more precisely for the ordinal preference/non-transferable utility environment. A few alternative concepts will be considered in

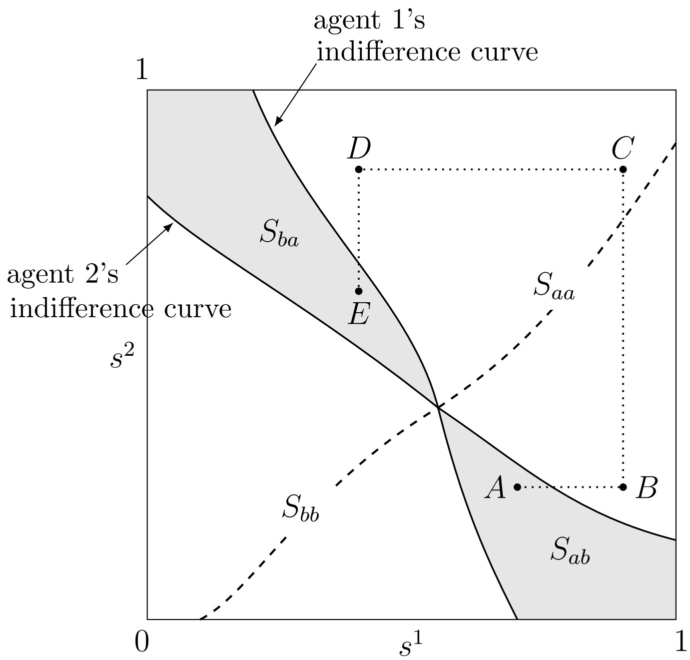

Figure 1: Impossibility of Ex-Post Incentive Compatibility

is a direct mechanism that is ex post incentive compatible and Pareto efficient. Then, at state $A = (s^1_A, s^2_A) \in S_{ab}$, agent 1 must receive a and agent 2 must receive b, and reporting the signal truthfully is a mutual best response. Now consider state $B = (s^1_B, s^2_A)$. Note that B differs from A only in agent 1's signal, and further that agent 1's (ordinal) preference remains unchanged. These two facts mean that, for the mechanism to be ex post incentive compatible, agent 1 must receive a at B; or else, agent 1 has incentives to report s^1_A instead and receive a for sure. Hence, the assignment remains unchanged between states A and B. Now consider state $C = (s^1_B, s^2_C)$. State C differs from state B only in agent 2's signal, and that agent's preference is the same between B and C. This means that the allocation must be the same at these two states. To see this, suppose for contradiction that agent 2 receives a with positive probability at C (and b with the remaining probability, as required by Pareto efficiency). Then agent 2 has incentives to misreport her signal at state B, reporting s^2_C instead. The same logic implies that the allocation at state $D = (s^1_D, s^2_C)$ must be exactly the same as the allocation at C. And similarly, the allocations at $E = (s^1_D, s^2_E)$

the paper, but the distinction in the notions does not matter here since there are only two objects.

must be the same as the allocation at D. Recalling the series of equivalences, we conclude that the allocation at E must be the same as the one at state A — that is, agent 1 receives a and agent 2 receives b. But this allocation is not Pareto efficient since $E \in S_{ba}$, showing that there exists no mechanism that is Pareto efficient and ex post incentive compatible.

In fact, the above argument implies much more than merely the impossibility of ex post incentive compatibility and Pareto efficiency. We can show that any ex post incentive compatible mechanism (that assigns both objects) can only implement a trivial allocation that is constant across all states! To see this, take two states arbitrarily, say s and $\hat{s} \neq s$. As above, one can construct a connected path of states, $s_0 \rightarrow s_1 \rightarrow .... \rightarrow s_m$, such that $s_0 = s, s_m = \hat{s}$, any adjacent states s_j and s_{j+1} have the same signal for one agent and different signals for the other, and the latter agent's ordinal preferences are the same and strict for both states. Then ex post incentive compatibility implies that the allocation is unchanged between any adjacent states.

Several remarks are worth making. First, the latter "triviality" result—that the ex post incentive compatibility means that only a constant allocation can be implemented—is reminiscent of Jehiel et al. (2006), who arrive at the same conclusion *under the transferable utility* setup. Despite the resemblance, however, the current result is not implied by theirs. One reason is that their result requires that agents have multi-dimensional signals while ours does not. In fact, the absence of monetary transfers is needed for our result. If monetary transfers were available in our example, a Pareto efficient and ex post incentive compatible mechanism exists, despite the fact that Pareto efficiency would entail a stronger allocative requirement in the presence of monetary transfer.

This can be seen as follows. Note first that, given transferable utilities, Pareto efficiency implies **utilitarian efficiency**, which requires that agents 1 and 2 receive a and b, respectively, if $u^1(s) > u^2(s)$ (which corresponds to the region below the dashed curve in Figure 1) and b and a, respectively, if $u^1(s) < u^2(s)$ (the region above the dashed curve in Figure 1). To see how this outcome can be implemented in an ex post incentive compatible mechanism, let

$$\sigma^i(s^j) := \sup\{s^i \in [0,1] | u^j(s^1, s^2) \geq u^i(s^1, s^2)\}$$

for $i, j = 1, 2, i \neq j$, if the set is nonempty, or else let $\sigma^i(s^j) := 0$. Suppose that the mechanism designer collects reports $(s^1, s^2) \in [0,1]^2$ from the agents and assigns the objects in a utilitarian-efficient manner, while charging agent i a tariff $p^i(s^j) := u^i(\sigma^i(s^j), s^j)$ whenever she receives a. It then follows that agent i prefers a to b at any signal profile in which she receives a and that she prefers b to a at any signal profile in which she receives b, so truthful reporting is ex post incentive compatible.[12]

Second, the particular assumptions made above — convexity of $S_{o_1o_2}$, the single crossing property, and the assumption that there are only two agents and two objects — are not needed for the impossibility of implementing the efficient allocation above. Section 4 will establish inefficiency in a general setting in which these assumptions are relaxed. By contrast, the second impossibility result above (i.e. the impossibility of implementing nontrivial allocations) does not generalize straightforwardly beyond the two-agents two-objects case. To see this, suppose that there are three agents, 1, 2, and 3, and three objects, a, b, and c. A mechanism that always assigns object c to agent 3, but assigns a and b between agents 1 and 2 in a way that varies only with agent 3's signal, is clearly ex post incentive compatible. This example points to another difference of the current model from Jehiel et al. (2006). Unlike their public decision problem, agents are indifferent across some allocations in our setting; for instance, agent 3 is indifferent on how a and b are allocated between 1 and 2. This indifference can be exploited to implement a nontrivial allocation in ex post incentive compatible mechanisms. In Subsection 4.3, we provide a generalization of the second impossibility result by strengthening the incentive requirement to ex post group incentive compatibility.

Third, the impossibility results without monetary transfers rest crucially on ex post incentive compatibility. Interdependence of preferences does not preclude efficiency if one

[12]The detailed argument for incentive compatibility is similar to that in Maskin (1992) and as follows. By the single crossing property and the definitions of σ^1 and σ^2, if $u^i(s) - u^j(s) \geq 0$, then $s^i \geq \sigma^i(s^j)$ and $s^j \leq \sigma^j(s^i)$, and if $u^i(s) - u^j(s) = 0$, then $s^i = \sigma^i(s^j)$ and $s^j = \sigma^j(s^i)$. Suppose first $u^1(s) - u^2(s) \geq 0$. Then, $s^1 \geq \sigma^1(s^2)$, so

$$u^1(s^1, s^2) - p^1(s^2) = u^1(s^1, s^2) - u^1(\sigma^1(s^2), s^2) \geq 0.$$

Hence, she (weakly) prefers a to b, so she will have incentives to report her signal truthfully. Suppose $u^1(s) - u^2(s) \leq 0$. Then $s^1 \leq \sigma^1(s^2)$, so

$$u^1(s^1, s^2) - p^1(s^2) = u^1(s^1, s^2) - u^1(\sigma^1(s^2), s^2) \leq 0.$$

Hence, the agent again has incentives to report her signal truthfully. The argument is symmetric for agent 2.

relaxes the equilibrium notion to Bayesian Nash equilibrium. Perhaps surprisingly, it is possible to implement a Pareto efficient allocation, even without transfers, in a Bayesian incentive compatible mechanism, when the agents' preferences are sufficiently congruent. We will present this Bayesian possibility result in Section 5.

4 Ex Post Incentive Compatible Mechanisms

4.1 Setup

Suppose there are n agents N and n objects O. A (pure) **assignment** is a one-to-one mapping μ from N to O, where $\mu^i = a$ means that agent i is assigned object a under assignment μ. Let $\mathcal{M}$ be the set of all assignments. A **random assignment** is a probability distribution over pure assignments. A random assignment $P \in \Delta(\mathcal{M})$[13] assigns agent $i \in N$ to object $a \in O$ with probability

$$P_a^i = \sum_{\mu \in \mathcal{M}} P(\mu) \mathbf{1}_{\{\mu^i = a\}},$$

where $\mathbf{1}_{\{\mu^i = a\}}$ is the indicator function (whose value is one if $\mu^i = a$ and zero otherwise).

Each agent i receives a private signal $s^i \in S^i$. We denote a profile of signals by $s = (s^1, ..., s^n) \in S \equiv \prod_{i \in N} S^i$. We assume that S^i is a convex subset of R^{m^i} for some nonnegative integer m^i for each $i \in N$. Agent i has value $v_a^i(s)$ for object a at signal profile s, which is differentiable (thus continuous in particular) in s. Let $\pi : O \to \{1, ..., n\}$ be a function that represents ordinal preferences of an individual: An agent with preference π prefers a to b if $\pi_a < \pi_b$ and is indifferent between them if $\pi_a = \pi_b$. For each agent i and signal profile s, agent i's value function v^i induces an associated preference relation $\pi^i(s)$ where $\pi_a^i(s)$ denotes the ranking of object a in preference relation $\pi^i(s)$ induced by the value function $v^i(s)$: Formally, $\pi_a^i(s) < \pi_b^i(s)$ if and only if $v_a^i(s) > v_b^i(s)$. A preference relation π^i is said to be **strict** if $\pi_a^i \neq \pi_b^i$ for any pair of objects $a \neq b$. A preference profile $\pi = (\pi^i)_{i \in N}$ is said to be strict if π^i is strict for every $i \in N$.

An assignment μ is **Pareto efficient** at preference profile π if there exists no assignment $\hat{\mu}$ such that $\pi_{\hat{\mu}^i}^i \leq \pi_{\mu^i}^i$ for all $i \in N$, with strict inequality for at least one $i \in N$. A

[13]Given set X, we denote by $\Delta(X)$ a probability distribution over X.

mechanism is a mapping $\varphi : S \to \Delta(\mathcal{M})$ from a vector $s \in S$ of signals to a random assignment. A mechanism φ is **Pareto efficient** if, for all $s \in S$, every (pure) assignment in the support of $\varphi(s)$ is Pareto efficient at $\pi(s) = (\pi^i(s))_{i \in N}$. In other words, we focus on "ex post" Pareto efficiency, although we will omit "ex post" since we do not consider any other efficiency notion in this paper.

We now introduce incentive compatibility concepts we shall use. To begin, we say that a random assignment P **first-order stochastically dominates** another random assignment $\hat{P}$ for i at preference π^i if

$$\sum_{b \in O : \pi_b^i \leq \pi_a^i} P_b^i \geq \sum_{b \in O : \pi_b^i \leq \pi_a^i} \hat{P}_b^i,$$

for all $a \in O$. If all these inequalities hold and at least one of them holds strictly, then we say that P **strictly first order stochastically dominates** $\hat{P}$ at π^i. We then say that a mechanism φ is **weakly ex post incentive compatible** if there exist no agent i, signal profile $s = (s^i, s^{-i}) \in S$, and signal $\bar{s}^i$ for i such that $\varphi(\bar{s}^i, s^{-i})$ strictly first-order stochastically dominates $\varphi(s^i, s^{-i})$ at $\pi^i(s)$. A mechanism is **ex post incentive compatible** if, for every agent i and signal profile $s = (s^i, s^{-i})$, $\varphi(s^i, s^{-i})$ first-order stochastically dominates $\varphi(\bar{s}^i, s^{-i})$ for all $\bar{s}^i$ at $\pi^i(s)$. Unlike weak ex post incentive compatibility, ex post incentive compatibility even eliminates the possibility that an agent's random assignments under true and misreported preferences are incomparable with respect to first-order stochastic dominance. Clearly, ex post incentive compatibility is a stronger requirement than weak ex post incentive compatibility.

Remark 1. As is clear, our notions of incentive compatibility are ordinal.[14] The ordinal concept has the additional benefit of being robust to the specific assumptions about agents' attitudes toward risk or uncertainty. One could alternatively define ex post incentive compatibility based on expected utilities. This alternative concept is weaker than our notion of ex post incentive compatibility but stronger than weak ex post incentive compatibility. Note that all these concepts coincide if one restricts attention to deterministic mechanisms as is often done in mechanism design, for instance Jehiel et al. (2006).

[14] Bogomolnaia and Moulin (2001) define ordinal incentive compatibility concepts in private-values environments. Our concepts reduce to theirs under private values.

4.2 Inefficiency in Weak Ex Post Implementation

We now present our first impossibility result. To do so, we introduce a number of assumptions on the signal space. The first assumption is central to our study: it formalizes the requirement that there be at least some interdependence in agents' valuations.

Assumption 1 (Interdependence)**.** *For any $i, j \in N$, $a, b \in O$ such that $a \neq b$, and $s \in S$ such that $v_a^i(s) = v_b^i(s)$, there exists $z^j \in \mathbb{R}^{m^j}$ with $\|z^j\| = 1$ such that $\nabla_{z^j} v_a^i(s) \neq \nabla_{z^j} v_b^i(s)$.*[15]

This assumption requires that agent j's signal influences agent i's relative preferences between any pair of objects, at least when agent i is indifferent between these two objects. This condition captures the notion of interdependence. It is worth noting that the condition does not require the value interdependence to be large. As will be seen from Example 1, the condition may hold even with very little interdependence (i.e., almost private values).

The next assumption means that the signal space is sufficiently rich. To state the condition, fix any pair of agents i and j, two objects a and b, and signal profile $s^{-ij} \in S^{-ij}$. Then, for $k, k' \in \{a, b\}$, we define $S_{kk'}^{ij}(s^{-ij}) \subset S^i \times S^j$ to be the (open) set of i and j's signal profiles for which (i) agent i ranks k above $o \in \{a, b\} \backslash \{k\}$, and o above any $k'' \notin \{a, b\}$, (ii) agent j ranks k' above $o \in \{a, b\} \backslash \{k'\}$, and o above any $k'' \notin \{a, b\}$, (iii) all other agents rank both a and b below any $k'' \notin \{a, b\}$ (we suppress the dependence of $S_{kk'}^{ij}(s^{-ij})$ on the set of objects $\{a, b\}$ to simplify notation).

Assumption 2 (Rich Domain)**.** *There exist $i, j \in N$, $a, b \in O$, and $s^{-ij} \in S^{-ij}$ such that $S_{kk'}^{ij}(s^{-ij})$ is non-empty for all $k, k' \in \{a, b\}$.*

As suggested by the name, the Rich Domain assumption postulates that the signal structure is rich enough to represent various preference profiles. Specifically, it means that one should be able to find two agents, a signal profile for all other agents, and two objects a and b such that the two prefer a and b to the other objects, the other agents find them

[15] Here, $\nabla_{z^j} v_o^i(s), o = a, b$, denotes the directional derivative of the function v_o^i along a given vector $z^j \in \mathbb{R}^{m^j}$ with Euclidean norm $\|z^j\| = 1$ at a given signal profile $s \in S$. To be concrete, $\nabla_{z^j} f(s) := \lim_{h \to 0} \frac{f(s^j + hz^j, s^{-j}) - f(s^j, s^{-j})}{h}$ for any $z^j \in \mathbb{R}^{m^j}$ with $\|z^j\| = 1$ such that $s^j + hz^j \in S^j$ for sufficiently small $h > 0$.

the two least preferred, and the two agents find either a or b to be the most preferred depending on their signals.

In order to state the next assumption, fix the two agents i and j, two objects a and b, and signal profile $s^{-ij} \in S^{-ij}$ as before. For $k \in \{a, b\}$, let $S^{ij}_{k\bullet}(s^{-ij}) \subset int(S^i \times S^j)$ denote the set of i and j's interior signal profiles[16] for which (i) and (iii) in the above definition of $S^{ij}_{kk'}(s^{-ij})$ hold but the property (ii) is relaxed to: (ii') agent j ranks a and b above any $k'' \notin \{a, b\}$. That is, $S^{ij}_{k\bullet}(s^{-ij})$ differs from $S^{ij}_{kk'}(s^{-ij})$ in that agent j's ranking between a and b is unspecified in the former set. Similarly, $S^{ij}_{\bullet k'}(s^{-ij})$ denotes the set of i and j's interior signal profiles for which (ii) and (iii) in the above definition of $S^{ij}_{kk'}(s^{-ij})$ hold but (i) is replaced by a weaker property: (i') agent i ranks a and b above any $k'' \notin \{a, b\}$.

Assumption 3 (Connectedness)**.** *For some $i, j \in N$, $a, b \in O$, and $s^{-ij} \in S^{-ij}$ satisfying the Rich Domain assumption (Assumption 2), and for some $k \in \{a, b\}$, both $S^{ij}_{k\bullet}(s^{-ij})$ and $S^{ij}_{\bullet k}(s^{-ij})$ are connected.*[17]

In our context of Euclidean spaces, connectedness of the open set $S^{ij}_{k\bullet}(s^{-ij})$ means that any two points in that set can be linked by a path contained in that set. Roughly, it means that the ordinal preferences vary stably with the changes in signals of agent i and j when others' signals remain fixed. This condition is relatively mild and in particular weaker than the assumption that the set $S^{ij}_{k\bullet}(s^{-ij})$ is convex.

We are now ready to present our first impossibility theorem (all proofs are in the Appendix A).

Theorem 1. *Under the assumptions of Interdependence, Rich Domain, and Connectedness, there exists no mechanism φ that is both Pareto efficient and weakly ex post incentive compatible.*

The key assumption used for the theorem is Rich Domain. Clearly, this assumption is easier to satisfy when each agent's signal is multidimensional, but multidimensionality of individual signals is not needed for the assumption. In fact, the Rich Domain assumption

[16] By taking the interior of $S^i \times S^j$, we are ruling out points on the boundary of $S^i \times S^j$, which will give us some room to perturb signal profiles when needed.

[17] Note that a set is **connected** if it cannot be partitioned into two sets that are open in the relative topology.

is satisfied even in fairly natural models with single dimensional signals. This point is illustrated in the following example, which one can see as a natural extension of the two-agent example described in the earlier Section 3.[18]

Example 1 (Canonical one-dimensional signal model). Assume that each agent i has signal $s^i \in [0,1]$. The set of objects is given by $O = \{o_1, ..., o_n\}$. Given signal profile $s \in [0,1]^n$, agent i's utility from object o_k is given by $v^i_{o_k}(s) = \alpha_k w^i(s) + \beta_k$, where $\alpha_k, \beta_k > 0$ and $w^i(s) := \gamma s^i + (1-\gamma)\frac{\sum_{j\neq i} s^j}{n-1}$ with $\gamma \in (\frac{1}{2}, 1)$. Assume that $\beta_{k-1} - \beta_k = \delta$ for some $\delta > 0$ and $\Delta_k := \alpha_k - \alpha_{k-1}$ is positive and strictly decreasing in k for $k = 2, ..., n$. We assume in addition that $\Delta_n \gamma > \delta > \Delta_2(1-\gamma)$.[19] The first (resp. second) inequality, combined with the previous assumptions, implies that if s^i is sufficiently close to 1 (resp. 0), then agent i prefers o_n the most (resp. least) and o_{n-1} the second most (resp. second least) irrespective of the others' signals.[20] This is illustrated in Figure 2 below for the case of 3 agents and 3 objects. As can be seen in the figure, for $s^1, s^2 \simeq 1$ and $s^3 \simeq 0$, agents 1 and 2 prefer $o_3 - o_2 - o_1$ in that order, whereas agent 3 prefers $o_1 - o_2 - o_3$ in that order.

Let $a = o_n$ and $b = o_{n-1}$, and fix each s^k, $k \neq i, j$, to be sufficiently close to zero so that a and b are the two least preferred objects for agent $k \neq i, j$, irrespective of i and j's signals. Now, one can find a signal profile $(\hat{s}^i, \hat{s}^j) \in (0,1)^2$ for agents i and j such that

$$w^i(\hat{s}^i, \hat{s}^j, s^{-ij}) = w^j(\hat{s}^i, \hat{s}^j, s^{-ij}) = \frac{\delta}{\Delta_n},$$

which means that given the signal profile $\hat{s} = (\hat{s}^i, \hat{s}^j, s^{-ij})$, both agents i and j are indifferent between a and b.[21] It is easy to check that this condition also implies that both i and j prefer

[18] While the utility functions are linear in the example, the linearity assumption is made only for convenience. Our assumptions of Interdependence, Rich Domain, and Connectedness can be seen to hold with nonlinear utility functions which possess the same qualitative features as the linear utility functions described here.

[19] Since $\Delta_2 > \Delta_n$, this assumption requires γ to be sufficiently large. One could think of this as a strengthening of the single crossing property. As long as this assumption is satisfied, we can allow for asymmetric value functions with α^k, β^k, and γ differing across agents.

[20] To see this, we can obtain $v^i_{o_k}(s) - v^i_{o_{k-1}}(s) = \Delta_k(\gamma s^i + (1-\gamma)\frac{\sum_{j\neq i} s^j}{n-1}) - \delta$ and note that this expression is positive (resp. negative) for all k and s^{-i} if $\Delta_n\gamma > \delta$ and $s^i \simeq 1$ (resp. $\Delta_2(1-\gamma) < \delta$ and $s^i \simeq 0$).

[21] To see that such $\hat{s}^i$ and $\hat{s}^j$ exist, first take $s^i = s^j$, which implies $w^i(s^i, s^j, s^{-ij}) = w^j(s^i, s^j, s^{-ij})$ by definition of $w^i(\cdot)$ and $w^j(\cdot)$. Then note that these are smaller than δ/Δ_n for $s^i = s^j = 0$, while the reverse inequality holds for $s^i = s^j = 1$. Since utility functions are continuous, by the mean value theorem there exists a value $\hat{s}^i = \hat{s}^j \in (0,1)$ of the signals such that the desired equality holds.

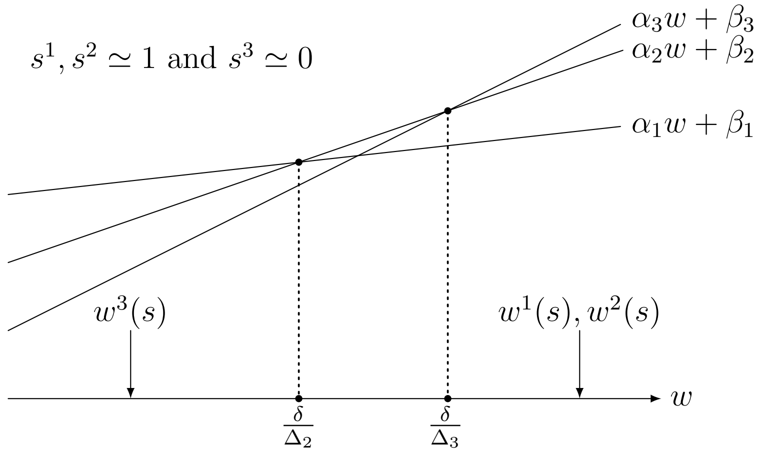

Figure 2: Illustration of Rich Domain Assumption with $n = 3$

a and b to all other objects. Then, the Rich Domain assumption is satisfied with small $\varepsilon > 0$ chosen so that $(\hat{s}^i+\varepsilon, \hat{s}^j+\varepsilon) \in S^{ij}_{aa}(s^{-ij})$, $(\hat{s}^i+\varepsilon, \hat{s}^j-\varepsilon) \in S^{ij}_{ab}(s^{-ij})$, $(\hat{s}^i-\varepsilon, \hat{s}^j+\varepsilon) \in S^{ij}_{ba}(s^{-ij})$, and $(\hat{s}^i - \varepsilon, \hat{s}^j - \varepsilon) \in S^{ij}_{bb}(s^{-ij})$.

Further, the Connectedness assumption is satisfied since the linearity of $v^i_{o_k}$ means that $S^{ij}_{k\cdot}(s^{-ij})$ and $S^{ij}_{\cdot k}(s^{-ij})$, each of which is a set defined by finitely many linear inequalities, are convex. Finally, to see that the Interdependence (Assumption 1) condition holds, suppose that agent i is indifferent between objects o_k and o_ℓ where $k > \ell$. It then follows that $(\alpha_k - \alpha_\ell)w^i(s) = \beta_\ell - \beta_k$. It is easy to see that this tie is broken by a slight change in any agent's signal since $\frac{\partial w^i(s)}{\partial s^j} > 0, \forall i, j$. In particular, the required interdependence $(1-\gamma) > 0$ can be arbitrarily small, in which case the agents' preferences become almost private.

Remark 2. As stated in Remark 1, our (ordinal) notion of weak ex post incentive compatibility is weaker than the cardinal notion of ex post incentive compatibility based on expected utilities. Hence, our inefficiency result continues to hold when one employs the latter concept of incentive compatibility.

4.3 Limits of Ex Post Group Incentive Compatibility

In this section, we consider joint manipulations by multiple agents, and a mechanism that is robust against such manipulations in the ex post sense. Formally, we say that a mechanism φ is **manipulable by group** $N' \subset N$ **at** $s \in S$ if there exists a signal profile $\hat{s}^{N'} \in \prod_{i\in N'} S^i$

such that, for all $i \in N'$, $\varphi(s^{N'}, s^{-N'})$ does not strictly first-order stochastically dominate $\varphi(\hat{s}^{N'}, s^{-N'})$ at $\pi^i(s)$, and $\varphi^i(s^{N'}, s^{-N'}) \neq \varphi^i(\hat{s}^{N'}, s^{-N'})$ for at least one $i \in N'$. A mechanism φ is said to be **ex post group incentive compatible** if it is not manipulable by any group $N' \subset N$ at any $s \in S$. This concept is a strengthening of ex post incentive compatibility, requiring that the mechanism eliminates profitable misreporting of preferences not only by an individual agent, but also by a group of agents. We will show that this strengthening of incentive compatibility leads to an even stronger impossibility result, namely that only a constant allocation can be implemented.

To begin, we say that a mechanism φ is **trivial** if $\varphi(s) = \varphi(\hat{s})$ for all $s, \hat{s} \in S$ such that $\pi(s)$ and $\pi(\hat{s})$ are strict preference profiles. Our result is that any ex post group incentive compatible mechanism must be trivial. To obtain this result, we shall invoke again the Interdependence assumption (Assumption 1), and two variants of Assumptions 2 and 3:[22]

Assumption 4 (Rich Domain*)**.** *For any preference profile π, there exists a signal profile $s \in int(S)$ such that $\pi(s) = \pi$.*

The Rich Domain* assumption requires that, given any preference profile, there is an interior signal that induces it.

Assumption 5 (Connectedness*)**.** *For any strict preference profile π, the set $S_\pi := \{s \in S | \pi(s) = \pi\}$ is connected.*

Theorem 2. *Under the assumptions of Interdependence, Rich Domain*, and Connectedness*, if φ is ex post group incentive compatible, then φ is trivial.*

To see why group incentive compatibility is needed for this result, recall the example in Section 3 with three agents, 1, 2, and 3, and three objects, a, b, and c. Consider a mechanism that always assigns object c to agent 3, but assigns a and b between agents 1 and 2 in a way that varies only with the signal of agent 3. Such a mechanism is ex post incentive compatible, but it is not ex post group incentive compatible since either 1 or 2 will stand to gain from a joint manipulation with agent 3.[23]

[22] It is straightforward, if tedious, to verify that there is no logical relationship between the Rich Domain and Rich Domain* assumptions or between the Connectedness and Connectedness* assumptions. See Appendix B.

[23] To see why this is the case, first note that by the Rich Domain* condition, there always exists a signal

While ex post group incentive compatibility is a strong requirement, the triviality result is not expected from the traditional private value model. Observe that when the values are private, our notion of ex post group incentive compatibility reduces to group strategy-proofness (see Papai (2000) and Pycia and Ünver (2009) for instance). This latter requirement is met by a large class of mechanisms that attain efficiency in the private value setting (Pycia and Ünver, 2009). In this regard, the triviality result of Theorem 2 is striking, particularly since it holds even when the preferences are "almost" private.

To obtain the intuition of the proof, let s and $\hat{s}$ be signal profiles at which preferences of all agents are strict. We construct a connected path of states, $s_0 \to s_1 \to .... \to s_m$, such that $s_0 = s, s_m = \hat{s}$, preferences of all agents are strict at each of these states, and

- any adjacent states s_k and s_{k+1} differ in the signal of only one agent, say j_k, and
- the ordinal preferences remain unchanged between s_k and s_{k+1} for all agents except for at most one agent, say i_k, who is different from j_k.

Between two adjacent states s_k and s_{k+1}, the assignment for agent j_k (whose signal varies across those states) cannot change due to ex post incentive compatibility (since her ordinal preferences are strict and remain unchanged per our construction). Ex post group incentive compatibility then implies that the assignments for every other agent whose preferences do not change should remain unchanged as well, because otherwise the agent whose assignment changes can profitably manipulate jointly with j_k. It then follows that the assignment for i_k (whose strict preferences vary across the states as described above) must also remain unchanged, since the assignments for all other agents remain unchanged (recall that there exists at most only one agent, i_k, whose preferences vary, by our construction). Thus the entire assignments remain unchanged between s_k and s_{k+1} for each k, and hence between s and $\hat{s}$, which implies the result. The detailed proof is in the Appendix.

While Assumptions 1, 4, and 5 enable us to construct a connected path of states with the above desired properties, the existence of such a path does not require the full force

profile $s = (s^1, s^2, s^3)$ such that both agents 1 and 2 prefer a to b. By definition of the mechanism, there exists signal $\hat{s}^3$ such that $\varphi(s^1, s^2, s^3) \neq \varphi(s^1, s^2, \hat{s}^3)$. Since agent 3 receives c with certainty at any signal profile, $\varphi^i_a(s^1, s^2, s^3) < \varphi^i_a(s^1, s^2, \hat{s}^3)$ for an agent $i \in \{1, 2\}$. Thus, agent 3 can report $\hat{s}^3$ to benefit agent i at s, so φ is manipulable by $\{i, 3\}$ at s.

of Rich Domain* and Connectedness*, and our proof method can be applied even to cases without those assumptions. To see this, recall the canonical one-dimensional signal model in Example 1. The preference specification of this example does not admit a full set of ordinal preferences, so it does not satisfy Rich Domain*.[24] Yet, it can be shown that there exists a connected path required for the proof of the theorem.

Proposition 1. *Any ex post group incentive compatible mechanism is trivial in the Canonical one-dimensional signal model in Example 1.*

Remark 3. The "non-wastefulness" feature of the house allocation model — that all objects are assigned — plays a role in the constancy result of Theorem 2. Without this assumption, it is ex post group incentive compatible, for instance, to assign agent 1 his most preferred object (which varies across signals) and assign all other agents "no" objects; the agents would then have no incentive to lie about signals individually or jointly. In this sense, Theorem 2 can be rephrased as establishing "constancy" among non-wasteful mechanisms. As can be inferred from the example, though, there is a sense in which the extent of implementable "variation" in allocation is limited even in a wasteful mechanism. If one restricts attention to a deterministic mechanism (one that implements a deterministic allocation for each profile of signals), then only one agent's allocation can change between any two signal profiles.[25]

Remark 4. Ex post group incentive compatibility is a strong requirement. However, a closer look at the proof reveals that the full force of this condition is *not* needed for the result. More specifically, the only requirement we need is that *no individual or pair of agents* can benefit from misreporting their preferences. In other words, precluding manipulations by groups of arbitrary sizes is not needed. To see this point, simply observe that the proof, as outlined above, applies the condition for only individuals and pairs.

[24]This can be seen easily for the case with $n = 3$. Letting $a := o_3, b := o_2, c := o_1$, the specification admits preference orderings: abc, bac, bca, cba, but it does not admit acb or cab.

[25]Roughly speaking, if allocations vary for two agents (call them "inside" agents), even only as a function of the signals of the other agents (call them "outside" agents), there is a scope for the outside agents to jointly manipulate with one of the inside agents to improve his assignment, when both inside agents prefer the same object.

5 Bayesian Incentive Compatible Mechanisms

In this section, we relax the incentive requirement by considering mechanisms that support truthful reporting as **Bayesian Nash equilibrium**. We show that the weakening of the incentive requirement enables us to achieve Pareto efficiency via a relatively simple mechanism, under some intuitive condition.

We focus here on the 2×2 case with single-dimensional signals. Suppose two objects a and b are assigned between agents 1 and 2. Each agent i's signal s^i is assumed to be drawn independently of s^j from the interval $[0, 1]$ via cdf F^i. As in Section 2, let $u^i(s) = v^i_a(s) - v^i_b(s)$ and assume $u^i(\cdot)$ to be increasing in both signals and satisfy the single crossing property (1).[26] A mechanism φ is **Bayesian incentive compatible** if truth-telling is a Bayesian Nash equilibrium of mechanism φ — namely, for each agent, reporting his true signal maximizes the expected utility.

In Figure 3, we reproduce the same indifference curves as in Figure 1, in which the two agents' indifference curves have a unique intersection.[27] We first observe that some well-known assignment mechanisms do not achieve Pareto efficiency. Consider for instance a **serial dictatorship** mechanism where agent 1 makes the first choice and agent 2 gets the remaining object. Without knowing agent 2's signal, he will choose a if $s^1 > \hat{s}^1$ and b if $s^1 < \hat{s}^1$, where $\hat{s}^1$ satisfies $\int_0^1 u^1(\hat{s}^1, s^2)dF^2(s^2) = 0$, i.e. agent 1 is indifferent between a and b in expectation at $\hat{s}^1$. Hence, agent 1 will end up with object a to the right of the dashed vertical line in the left panel of Figure 3, inefficiently obtaining a in the stroked area, unless $\hat{s}^1$ happens to coincide with $\bar{s}^1$. For the same reason, another well-known mechanism, **random serial dictatorship**, in which one agent is chosen at random to pick the preferred object, is inefficient unless the agents are symmetric. This will be seen shortly.

We propose an alternative mechanism, denoted φ^*, whose assignment probabilities are described in the right panel of Figure 3. The first number in the parenthesis represents the probability that agent 1 receives a, and the second number represents the probability

[26]Without this assumption, it is impossible to achieve Pareto efficiency even when transfers are allowed, as shown by Maskin (1992).

[27]Due to the single crossing condition, there can be at most one intersection. When there is no intersection, one can show that a Pareto efficient assignment is achieved through a trivial, constant, mechanism.

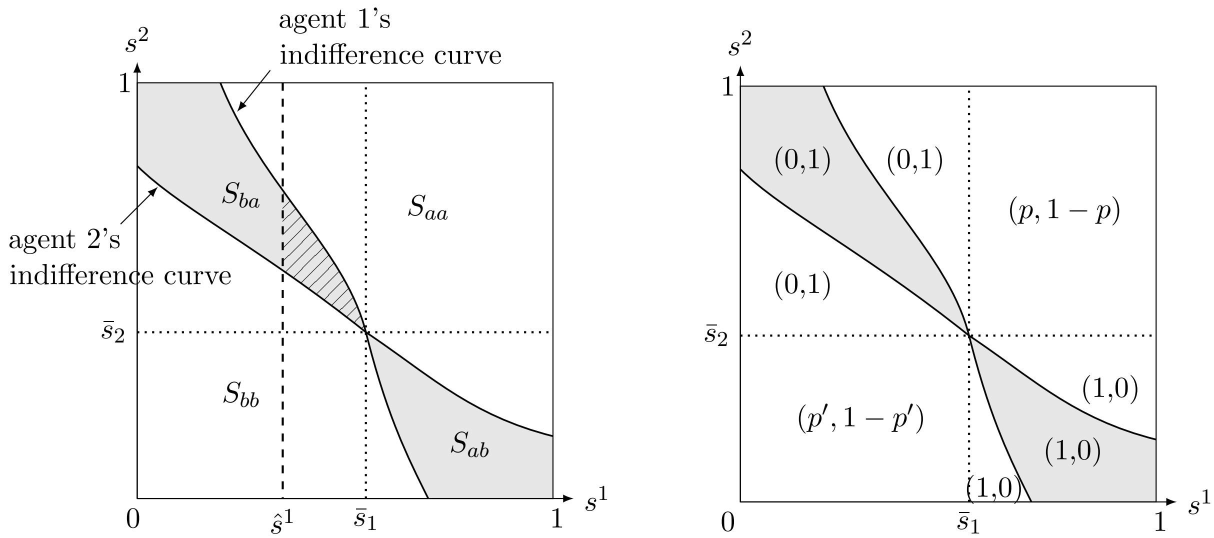

Figure 3: Bayesian Implementation

that agent 2 receives a. Clearly, this mechanism is Pareto efficient since agent 1 gets a (resp. b) in the area S_{ab} (resp. S_{ba}). The remaining question is whether one can find a pair $p, p' \in [0,1]$ that makes φ^* Bayesian incentive compatible. The following result says that such a pair exists if and only if the preferences of the two agents' threshold types are "congruent in expectation."

Theorem 3. *There exists a pair $p, p' \in [0,1]$ that makes φ^* Bayesian incentive compatible, if and only if either*

$$\int_0^1 u^1(\bar{s}^1, s^2)dF^2(s^2) \geq 0 \geq \int_0^1 u^2(s^1, \bar{s}^2)dF^1(s^1) \tag{2}$$

or

$$\int_0^1 u^1(\bar{s}^1, s^2)dF^2(s^2) \leq 0 \leq \int_0^1 u^2(s^1, \bar{s}^2)dF^1(s^1). \tag{3}$$

In fact, the mechanism φ^* can be implemented by modifying the serial dictatorship in the following way. Each agent reports (simultaneously) whether she prefers a or b. If the agents indicate they prefer different objects, they are assigned their preferred objects. If both agents indicate they prefer a, then agent 1 is chosen with probability p and agent 2 is chosen with the remaining probability to claim a. If both indicate they prefer b, then agent 1 is chosen with probability p' and agent 2 is chosen with the remaining probability to claim b.[28] With the probabilities p and p' given in Theorem 3, each agent i will indicate a to be the preferred object if and only if $s^i \geq \bar{s}^i$, provided that the other agent j does the same. Clearly, this equilibrium strategy will result in the same assignment probabilities as in the right panel of Figure 3. Note that this mechanism becomes equivalent to the random serial dictatorship if and only if $p = p' = 1/2$, which occurs only in nongeneric symmetric cases.

This result, together with Theorems 1 and 2, suggests that the two incentive requirements entail dramatic differences in what can be implemented at least for two agent cases. While efficient allocations can be implemented by a Bayesian incentive compatible mechanism, only a trivial constant allocation can be implemented if one insists upon ex post incentive compatibility. The difference remains relevant even in an "almost private value" model. For instance, consider a two-agent model in which $u^i(s) = \gamma s^i + (1-\gamma)s^{-i} - 0.5$, for $\gamma \in (1/2, 1)$ and s^i is drawn uniformly from $[0, 1]$. Then, the congruence assumption in the statement of Theorem 3 is satisfied regardless of $\gamma \in (1/2, 1)$, so an efficient assignment is Bayesian implementable. As γ goes to 1, the model approaches a pure private value model; yet the impossibility results under ex post implementation remain valid for all such $\gamma < 1$. By contrast, with private values (i.e. $\gamma = 1$), the efficient assignment is dominant strategy (and hence ex post) implementable so the added incentive requirement does not entail any loss.

The condition in Theorem 3 is sufficient for the existence of a Pareto efficient and Bayesian incentive compatible mechanism. Although the necessity of this condition for efficiency is unclear for general Bayesian mechanisms, we can at least show that the condition is also necessary when we restrict attention to **ex-post monotonic mechanisms**: that is, for each $i = 1, 2$ and for all s^j, $\varphi^i_a(\cdot, s^j)$ is non-decreasing.

Theorem 4. *There exists a Bayesian incentive compatible mechanism that is Pareto efficient and ex-post monotonic if and only if condition* (2) *or* (3) *holds.*

The following remarks discuss how our possibility result may be extended to a more general environment.

[28] Note that this mechanism is a proxy version in the sense that once each agent reports his message, then the mechanism assigns the objects according to the description on behalf of the agents.

Remark 5. Theorem 3 can be generalized to the case in which signals are (weakly) positively correlated. To be concrete, the same result can be proven with the assumption that, for each i and $j \neq i$, the conditional cdf $F^i(s^i|s^j)$ is nonincreasing in s^j, i.e. $F^i(\cdot|s^j)$ first-order stochastically dominates $F^i(\cdot|\hat{s}^j)$ for $\hat{s}^j < s^j$. The sufficient conditions in the statement are unchanged, except for replacing each cumulative distribution function $F^i(s^i)$ in the inequalities to a conditional cdf at the threshold type, $F^i(s^i|\bar{s}^j)$.

Remark 6. The method of constructing mechanism φ^* might be extended to the general $n \times n$ case. In other words, we may exploit the non-uniqueness of Pareto efficient allocations, by searching for mechanisms that randomize over Pareto efficient allocations whenever they are not unique. The degree of freedom in choosing the randomizations could then be utilized to generate the right incentives for the threshold types. As $n \geq 3$ gets large, however, the problem becomes complicated, since the number of threshold types increases exponentially and the degree of freedom for selecting randomizations becomes increasingly rich. We leave this extension for future research.

References

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

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

Abdulkadiroğlu, Atila, Parag A. Pathak, Alvin E. Roth, and Tayfun Sönmez. 2005. "The Boston Public School Match." *American Economic Review Papers and Proceedings*, 95: 368–372.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2005. "The New York City High School Match." *American Economic Review Papers and Proceedings*, 95: 364–367.

Abdulkadiroğlu, Atila, Parag A. Pathak, and Alvin E. Roth. 2009. "Strategy-proofness versus Efficiency in Matching with Indifferences: Redesigning the NYC High School Match." *American Economic Review*, 99(5): 1954–1978.

Balinski, Michel, and Tayfun Sönmez. 1999. "A tale of two mechanisms: student placement." *Journal of Economic Theory*, 84: 73–94.

Bergemann, Dirk, and Juuso Välimäki. 2002. "Information Acquisition and Efficient Mechanism Design." *Econometrica*, 70: 1007–1033.

Bikhchandani, Sushil. 2006. "Ex Post Implementation in Environments with Private Goods." *Theoretical Economics*, 1: 369–393.

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

Chakraborty, Archishman, Alessandro Citanna, and Michael Ostrovsky. 2010. "Two-sided Matching with Interdependent Values." *Journal of Economic Theory*, 145: 85–105.

Chakraborty, Archishman, and Alessandro Citanna. 2011. "Group Stability in Matching With Interdependent Values." mimeo.

Cremer, Jacques, and Richard McLean. 1985. "Optimal Selling Strategies under Uncertainty for a Discriminating Monopolist when Demands Are Interdependent." *Econometrica*, 53: 345–361.

Esö, Péter, and Eric Maskin. 2002. "Multi-Good Efficient Auctions with Multi-Dimensional Information." manuscript, University of Oxford.

Gale, David, and Lloyd S. Shapley. 1962. "College Admissions and the Stability of Marriage." *American Mathematical Monthly*, 69: 9–15.

Jehiel, P., and B. Moldovanu. 2006. "Allocative and informational externalities in auctions and related mechanisms." In *The Proceedings of the 9th World Congress of the Econometric Society.* , ed. Richard Blundell, Whitney Newey and Torsten Persson. Cambridge University Press.

Jehiel, Philippe, and Benny Moldovanu. 2001. "Efficient Design with Interdependent Valuations." *Econometrica*, 1237–59.

Jehiel, Philippe, Moritz Meyer ter Vehn, Benny Moldovanu, and William R. Zame. 2006. "The Limits of Ex-Post Implementation." *Econometrica*, 585–610.

Krishna, Vijay. 2003. "Asymmetric English Auctions." *Journal of Economic Theory*, 112: 261–288.

Maskin, Eric. 1992. "Auctions and Privatization." In *Privatization: Symposium in Honor of Herbert Giersch.* , ed. Horst Siebert. Tübingen:Mohr (Siebeck).

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

Pathak, P.A. 2011. "The mechanism design approach to student assignment." *Annual Review of Economics*, 3(1).

Perry, Motty, and Philip Reny. 2002. "An Efficient Auction." *Econometrica*, 70: 1199–1212.

Pycia, M., and M.U. Ünver. 2009. "A theory of house allocation and exchange mechanisms." *Boston College Working Papers in Economics.*

Roth, Alvin E. 1984. "The Evolution of the Labor Market for Medical Interns and Residents: A Case Study in Game Theory." *Journal of Political Economy*, 92: 991–1016.

Roth, Alvin E. 2002. "The economist as engineer: Game theory, experimentation, and computation as tools for design economics." *Econometrica*, 70: 1341–1378.

Roth, Alvin E., and Marilda A. O. Sotomayor. 1990. *Two-sided matching: a study in game-theoretic modeling and analysis.* Cambridge:Cambridge University Press.

Sönmez, Tayfun, and Uktu Ünver. 2009. "Matching, Allocation, and Exchange of Discrete Resources." forthcoming, Handbook of Social Economics, eds. Jess Benhabib, Alberto Bisin, and Matthew Jackson, Elsevier.

Svensson, Lars-Gunnar. 1999. "Strategy-proof Allocation of Indivisible Goods." *Social Choice and Welfare*, 16.

Appendix A: Proofs

In the proof of Theorem 1, we will invoke the following mathematical result.

Lemma 1 (Proposition 2.10, Stewart (1999))**.** *Suppose X is an open and connected subset of a (multi-dimensional) Euclidean space. Then X is "step-connected" in the following sense: For any $x, x' \in X$, there exists a sequence $x_0, x_1, \ldots, x_m$ in X such that*

1. *For each $k = 0, \ldots, m-1$, there exists $i(k) \in \{1, \ldots, n\}$ such that $x_k^{-i(k)} = x_{k+1}^{-i(k)}$[29];*
2. *$\overline{x_k x_{k+1}} \subset X, \forall k = 0, \ldots, m-1$ with $x_0 = x$ and $x_m = x'$, where $\overline{x_k x_{k+1}}$ denotes the line segment connecting x_k and x_{k+1} (including the end points).*

The set (or path) $\bigcup_{k=0}^{m-1} \overline{x_k x_{k+1}}$ in the above Lemma will be referred to as a **step-wise path between x and x'**.

Proof of Theorem 1. Assume for contradiction that φ is Pareto efficient and weakly ex post incentive compatible. Take any $i, j \in N$, $a, b \in O$, and $s^{-ij} \in S^{-ij}$ that satisfy the Rich Domain and Connectedness assumptions. Suppose without loss that the Connectedness assumption is satisfied with $k = a$. Note that $S^{ij}_{a\cdot}(s^{-ij})$ and $S^{ij}_{\cdot a}(s^{-ij})$ are step-connected since they are connected and open. We first observe that there exists a signal profile $\hat{s}^{ij} = (\hat{s}^i, \hat{s}^j) \in S^{ij}_{\cdot a}(s^{-ij})$ such that given $\hat{s} := (\hat{s}^{ij}, s^{-ij})$, for any $c \neq a, b$,

$$\begin{gathered} v^i_a(\hat{s}) = v^i_b(\hat{s}) > v^i_c(\hat{s}),\ v^j_a(\hat{s}) > v^j_b(\hat{s}) > v^j_c(\hat{s}),\ \textit{and} \\ v^k_c(\hat{s}) > \max\{v^k_a(\hat{s}), v^k_b(\hat{s})\}, \forall k \neq i, j. \end{gathered} \tag{4}$$

To see this, use the Rich Domain assumption to choose any $r^{ij} = (r^i, r^j) \in S^{ij}_{aa}(s^{-ij})$ and $t^{ij} = (t^i, t^j) \in S^{ij}_{ba}(s^{-ij})$. Then, by the Connectedness assumption, there must be some continuous path between r^{ij} and t^{ij} that is contained in $S^{ij}_{\cdot a}(s^{-ij})$. Given the continuity of that path and value functions, we must have some signal profile $\hat{s}^{ij} \in S^{ij}_{\cdot a}(s^{-ij})$ such that $v^i_a(\hat{s}^{ij}, s^{-ij}) = v^i_b(\hat{s}^{ij}, s^{-ij})$. Clearly, $\hat{s} = (\hat{s}^{ij}, s^{-ij})$ satisfies the desired preference relationship.

By the Interdependence assumption and relation (4), there exists $z^j \in \mathbb{R}^{m^j}$ such that $(\hat{s}^i, \hat{s}^j - z^j) \in S^{ij}_{ba}(s^{-ij})$ and $(\hat{s}^i, \hat{s}^j + z^j) \in S^{ij}_{aa}(s^{-ij})$. Since φ is Pareto efficient, $\varphi^j(\hat{s}^i, \hat{s}^j -$

[29] Note that $x^{-i} = (x^j)_{j \neq i}$ denotes the components of vector x except for its i'th components.

$z^j, s^{-ij}) = a$.[30] Since j strictly prefers a most at both $\pi^j(\hat{s}^i, \hat{s}^j - z^j, s^{-ij})$ and $\pi^j(\hat{s}^i, \hat{s}^j + z^j, s^{-ij})$ and φ is weakly ex post incentive compatible, we must have $\varphi^j(\hat{s}^i, \hat{s}^j + z^j, s^{-ij}) = a$. Again since φ is Pareto efficient, $\varphi^i(\hat{s}^i, \hat{s}^j + z^j, s^{-ij}) = b$. Therefore we conclude that there exists a signal profile $\tilde{s}^{ij} = (\tilde{s}^i, \tilde{s}^j) \in S^{ij}_{aa}(s^{-ij})$ such that $\varphi^i(\tilde{s}^{ij}, s^{-ij}) = b$ and $\varphi^j(\tilde{s}^{ij}, s^{-ij}) = a$ (simply define $\tilde{s}^i = \hat{s}^i$, $\tilde{s}^j = \hat{s}^j + z^j$).

Consider now any profile $\check{s}^{ij} \in S^{ij}_{ab}(s^{-ij}) \subset S^{ij}_{a\bullet}(s^{-ij})$. By Lemma 1, the connectedness of $S^{ij}_{a\bullet}(s^{-ij})$, and the fact that $\tilde{s}^{ij} \in S^{ij}_{aa}(s^{-ij}) \subset S^{ij}_{a\bullet}(s^{-ij})$, one can find a step-wise path $\bigcup_{k=0}^{m-1} \overline{s^{ij}_k s^{ij}_{k+1}} \subset S^{ij}_{a\bullet}(s^{-ij})$ between $\check{s}^{ij} = s^{ij}_0$ and $\tilde{s}^{ij} = s^{ij}_m$. Since this path as well as value functions are continuous, one can find some signal profile $\bar{s}^{ij} = (\bar{s}^i, \bar{s}^j) \in \overline{s^{ij}_\ell s^{ij}_{\ell+1}}$ for some ℓ such that (1) $v^j_a(\bar{s}^{ij}, s^{-ij}) = v^j_b(\bar{s}^{ij}, s^{-ij})$ and (2) $s^{ij} \in S^{ij}_{aa}(s^{-ij})$ for all profiles $s^{ij} \neq \bar{s}^{ij}$ on the (sub)step-wise path between $\bar{s}^{ij}$ and $\tilde{s}^{ij}$. (That is, $\bar{s}^{ij}$ is the last point on the step-wise path going from $\check{s}^{ij}$ to $\tilde{s}^{ij}$ at which agent j is indifferent between a and b.) One can also choose ℓ so that $\bar{s}^{ij} \neq s^{ij}_{\ell+1}$. We then prove the following claim:

Claim 1. *$\varphi^i(s^{ij}_k, s^{-ij}) = b$ and $\varphi^j(s^{ij}_k, s^{-ij}) = a$ for all $k = \ell + 1, \cdots, m$.*

Proof. Note first that the following is true:

1. $\varphi^i(s^{ij}_m, s^{-ij}) = \varphi^i(\tilde{s}^{ij}, s^{-ij}) = b$ and $\varphi^j(s^{ij}_m, s^{-ij}) = \varphi^j(\tilde{s}^{ij}, s^{-ij}) = a$,
2. For each $k = \ell + 1, \ldots, m - 1$, either $s^i_k = s^i_{k+1}$ or $s^j_k = s^j_{k+1}$,[31] and
3. $(s^i_k, s^j_k) \in S^{ij}_{aa}(s^{-ij})$ for $k = \ell + 1, \ldots, m$.

For each $k = \ell + 1, \ldots, m - 1$, by items 2 and 3 above and ex post weak incentive compatibility of φ, $\varphi^i(s^i_k, s^j_k, s^{-ij}) = \varphi^i(s^i_{k+1}, s^j_{k+1}, s^{-ij})$ if $s^i_k \neq s^i_{k+1}$ and $\varphi^j(s^i_k, s^j_k, s^{-ij}) = \varphi^j(s^i_{k+1}, s^j_{k+1}, s^{-ij})$ if $s^j_k \neq s^j_{k+1}$. In either case, this and the Pareto efficiency of φ imply $\varphi^h(s^i_k, s^j_k, s^{-ij}) = \varphi^h(s^i_{k+1}, s^j_{k+1}, s^{-ij})$ for $h = i, j$, which, combined with item 1 above, gives us the desired result. □

Now, to establish the desired contradiction, we consider two cases: (i) $\bar{s}^j = s^j_{\ell+1}$; (ii) $\bar{s}^i = s^i_{\ell+1}$.

[30] We write $P^j = a$ for a degenerate random assignment such that $P^j_a = 1$.

[31] This is because, by Lemma 1, we can take the sequence in such a way that s_k and s_{k+1} differ only in one dimension, and hence in one agent's signal.

Case (i): Let $\Delta s^i := s^i_{\ell+1} - \bar{s}^i$. Then, for all $\varepsilon \in (0,1]$, $(\bar{s}^i + \varepsilon\Delta s^i, \bar{s}^j) = (\bar{s}^i + \varepsilon\Delta s^i, s^j_{\ell+1}) \in \overline{\bar{s}^{ij} s^{ij}_{\ell+1}} \subset S^{ij}_{aa}(s^{-ij})$. With small enough $\varepsilon > 0$, it also holds that $(\bar{s}^i - \varepsilon\Delta s^i, s^j_{\ell+1}) \in S^{ij}_{ab}(s^{-ij})$. Note that $\varphi^i(s^{ij}_{\ell+1}, s^{-ij}) = b$ by the above Claim 1, and observe that since agent i likes a most at both $\pi^i(\bar{s}^i + \varepsilon\Delta s^i, s^j_{\ell+1}, s^{-ij})$ and $\pi^i(s^i_{\ell+1}, s^j_{\ell+1}, s^{-ij})$, the weak ex-post incentive compatibility and Pareto efficiency of φ imply $\varphi^i(\bar{s}^i + \varepsilon\Delta s^i, s^j_{\ell+1}, s^{-ij}) = \varphi^i(s^i_{\ell+1}, s^j_{\ell+1}, s^{-ij}) = b$. Similarly, the weak ex-post incentive compatibility for agent i also implies $\varphi^i(\bar{s}^i - \varepsilon\Delta s^i, s^j_{\ell+1}, s^{-ij}) = b$, which means the inefficiency arises since $(\bar{s}^i - \varepsilon\Delta s^i, s^j_{\ell+1}) \in S^{ij}_{ab}(s^{-ij})$.

Case (ii): By the Interdependence assumption and the fact in (1) above that $v^j_a(\bar{s}^{ij}, s^{-ij}) = v^j_b(\bar{s}^{ij}, s^{-ij})$, one can find $z^i \in \mathbb{R}^{m^i}$ such that $(\bar{s}^i + \varepsilon z^i, \bar{s}^j) \in S^{ij}_{aa}(s^{-ij})$ and $(\bar{s}^i - \varepsilon z^i, \bar{s}^j) \in S^{ij}_{ab}(s^{-ij})$ for all small enough $\varepsilon > 0$. The Pareto efficiency then implies that $\varphi^i(\bar{s}^i - \varepsilon z^i, \bar{s}^j, s^{-ij}) = a$, which, in turn, implies $\varphi^i(\bar{s}^i + \varepsilon z^i, \bar{s}^j, s^{-ij}) = a$ due to the weak ex-post incentive compatibility for agent i with $\bar{s}^i + \varepsilon z^i$. Thus, by the Pareto efficiency,

$$\varphi^j(\bar{s}^i + \varepsilon z^i, \bar{s}^j, s^{-ij}) = b. \tag{5}$$

With small enough ε, we also have $(\bar{s}^i + \varepsilon z^i, s^j_{\ell+1}) \in S^{ij}_{aa}(s^{-ij})$ since $(\bar{s}^i, s^j_{\ell+1}) = (s^i_{\ell+1}, s^j_{\ell+1}) \in S^{ij}_{aa}(s^{-ij})$. Then, since agent i likes a most at both $\pi^i(\bar{s}^i + \varepsilon z^i, s^j_{\ell+1}, s^{-ij})$ and $\pi^i(s^i_{\ell+1}, s^j_{\ell+1}, s^{-ij})$, the weak ex-post incentive compatibility and Pareto efficiency of φ imply $\varphi^i(\bar{s}^i + \varepsilon z^i, s^j_{\ell+1}, s^{-ij}) = \varphi^i(s^i_{\ell+1}, s^j_{\ell+1}, s^{-ij}) = b$, which, in turn, implies $\varphi^j(\bar{s}^i + \varepsilon z^i, s^j_{\ell+1}, s^{-ij}) = a$ by the Pareto efficiency of φ. Given this and (5), agent j with $\bar{s}^j$ would prefer (mis)reporting $s^j_{\ell+1}$ to obtain a rather than b when others' type profile is $(\bar{s}^i + \varepsilon z^i, s^{-ij})$. □

Proof of Theorem 2. Consider two signal profiles $s, \hat{s} \in S$ such that $\pi(s)$ and $\pi(\hat{s})$ are strict. We will assume $s, \hat{s} \in int(S)$ and show $\varphi(s) = \varphi(\hat{s})$. Later we will extend our argument to signals on the boundary.

Consider a sequence of strict preference profiles $\pi_0, \pi_1, \ldots, \pi_m$ and a sequence of non-strict preference profiles $\tilde{\pi}_0, \tilde{\pi}_1, \ldots, \tilde{\pi}_{m-1}$[32] such that

1. $\pi_0 = \pi(s)$, $\pi_m = \pi(\hat{s})$,

2. For each $k = 0, \ldots, m-1$, there exists $i_k \in N$ such that $\pi^{-i_k}_k = \tilde{\pi}^{-i_k}_k = \pi^{-i_k}_{k+1}$ and $\pi^{i_k}_k$, $\tilde{\pi}^{i_k}_k$, and $\pi^{i_k}_{k+1}$ are "adjacent" with one another where $\pi^{i_k}_k \neq \pi^{i_k}_{k+1}$ are strict while $\tilde{\pi}^{i_k}_k$ is non-strict. That is, there exist $a_k, b_k \in O$ such that $a_k \neq b_k$ whose rankings are (equal

[32] With abuse of notation, the subscript here does not specify objects (as in π_a and π_b) as before.

or) next to each other at $\pi^{i_k}_k$, $\tilde{\pi}^{i_k}_k$, and $\pi^{i_k}_{k+1}$ such that $\pi^{i_k}_k$, $\tilde{\pi}^{i_k}_k$, and $\pi^{i_k}_{k+1}$ are different only in the ranking between a_k and b_k and $\pi^{i_k}_{k,a_k} < \pi^{i_k}_{k,b_k}, \tilde{\pi}^{i_k}_{k,a_k} = \tilde{\pi}^{i_k}_{k,b_k}, \pi^{i_k}_{k+1,a_k} > \pi^{i_k}_{k+1,b_k}$.[33]

By the Rich Domain* assumption, there exists a sequence of signal profiles $s_0, s_1, \ldots, s_{m-1} \in int(S)$ such that $\pi(s_k) = \tilde{\pi}_k$ for each k. Take a sequence of agents $j_0, j_1, \ldots, j_{m-1}$ such that $j_k \neq i_k$ for each k (such agents exist because $|N| \geq 2$). By the Interdependence assumption and the fact that $s_0, s_1, \ldots, s_{m-1} \in int(S)$, there exist signals $(\tilde{s}^{j_k}_{k-}, \tilde{s}^{j_k}_{k+})_{k=0}^{m-1}$ such that $\pi(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \pi_k$ and $\pi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k) = \pi_{k+1}$.[34]

Claim 2. *For each $k = 0, 1, \ldots, m-1$,*

$$\varphi(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \varphi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k).$$

Proof. First note, by construction, that $\pi^{j_k}(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \pi^{j_k}(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k) = \pi^{j_k}_k$ and this preference is strict. These facts as well as (group) ex post incentive compatibility of φ imply that

$$\sum_{b:\pi^{j_k}_{k,b} \leq \pi^{j_k}_{k,a}} \varphi^{j_k}_b(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \sum_{b:\pi^{j_k}_{k,b} \leq \pi^{j_k}_{k,a}} \varphi^{j_k}_b(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k),$$

for each $a \in O$ (otherwise, group ex post incentive compatibility of φ is violated at either $(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k)$ or $(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k)$ by a singleton "group" j_k). These equalities imply

$$\varphi^{j_k}(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \varphi^{j_k}(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k). \tag{6}$$

To show $\varphi(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \varphi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k)$ suppose, for contradiction, that

$$\varphi^{j}(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) \neq \varphi^{j}(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k), \tag{7}$$

for some $j \in N$. By equality (6), $j \neq j_k$. If inequality (7) holds for i_k, then, by equality (6) and the assumption that each of the n objects are assigned to exactly one of the n agents, there is another agent $j \neq i_k$ for whom inequality (7) holds. Thus we can assume $j \neq i_k, j_k$ without loss of generality. Since $\pi^{j}(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) = \pi^{j}(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k) = \pi^{j}_k$ and this preference is

[33] For any agent i, index k, and object a, $\pi^i_{k,a}$ denotes the ranking of object a at preference π^i_k.

[34] The argument is similar to the one for constructing $\tilde{s}^i$ in the proof of Theorem 1.

strict for any such j by assumption, inequality (7) implies that there exists an object $a \in O$ such that

$$\sum_{b:\pi^j_{k,b} \leq \pi^j_{k,a}} \varphi^j_b(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) > \sum_{b:\pi^j_{k,b} \leq \pi^j_{k,a}} \varphi^j_b(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k), \text{ or}$$
$$\sum_{b:\pi^j_{k,b} \leq \pi^j_{k,a}} \varphi^j_b(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) < \sum_{b:\pi^j_{k,b} \leq \pi^j_{k,a}} \varphi^j_b(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k).$$

In the former case, φ is manipulable by $N' = \{j_k, j\}$ at $(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k)$ since $\varphi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k)$ does not strictly first-order stochastically dominate $\varphi(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k)$ for j_k, j and $\varphi^j(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k) \neq \varphi^j(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k)$. In the latter case, φ is manipulable by $N' = \{j_k, j\}$ at $(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k)$ since $\varphi(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k)$ does not strictly first-order stochastically dominate $\varphi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k)$ for j_k, j and $\varphi^j(\tilde{s}^{j_k}_{k-}, s^{-j_k}_k) \neq \varphi^j(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k)$. Therefore, φ is not ex post group incentive compatible. □

Claim 3. *For each k,*

$$\varphi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k) = \varphi(\tilde{s}^{j_{k+1}}_{(k+1)-}, s^{-j_{k+1}}_{k+1}).$$

Proof. First note that by construction of the signals,

$$\pi(\tilde{s}^{j_k}_{k+}, s^{-j_k}_k) = \pi(\tilde{s}^{j_{k+1}}_{(k+1)-}, s^{-j_{k+1}}_{k+1}) = \pi_{k+1}.$$

Also observe that $S_{\pi_{k+1}}$ is open because π_{k+1} is a strict preference profile and utility functions are continuous in signal profiles, and connected by the Connectedness* assumption. Thus, by Lemma 1, there exists a sequence $s(0), s(1), \ldots, s(\bar{l}) \in S_{\pi_{k+1}}$ and $i(0), i(1), \ldots, i(\bar{l}-1) \in N$ such that

1. $s(0) = (\tilde{s}^{j_k}_{k+}, s^{-j_k}_k), s(\bar{l}) = (\tilde{s}^{j_{k+1}}_{(k+1)-}, s^{-j_{k+1}}_{k+1})$,
2. $s(l)^{-i(l)} = s(l+1)^{-i(l)}$ for each $l \in \{0, 1, \ldots, \bar{l}-1\}$.

For each l, since $\pi(s(l)) = \pi_{k+1}$ is a strict preference profile and φ satisfies (group) ex post incentive compatibility, an argument analogous to what lead to relation (6) in the proof of Claim 2 implies

$$\varphi^{i(l)}(s(l)) = \varphi^{i(l)}(s(l+1)). \tag{8}$$

Since φ satisfies group ex post incentive compatibility and π_{k+1} is a strict preference profile, this implies $\varphi(s(l)) = \varphi(s(l+1))$ for each l by an argument similar to the last part of the proof of Claim 2. Thus $\varphi(s(0)) = \varphi(s(\bar{l}))$, completing the proof. □

To complete the proof of the Theorem, observe that Claims 2 and 3 imply that

$$\varphi(\tilde{s}_{0-}^{j_0}, s_0^{-j_0}) = \varphi(\tilde{s}_{(m-1)+}^{j_{m-1}}, s_{m-1}^{-j_{m-1}}). \tag{9}$$

By arguments identical to the proof of Claim 3,

$$\varphi(s) = \varphi(s_{0-}^{j_0}, s_0^{-j_0}), \tag{10}$$

$$\varphi(\hat{s}) = \varphi(s_{(m-1)+}^{j_{m-1}}, s_{m-1}^{-j_{m-1}}). \tag{11}$$

Relations (9)-(11) imply $\varphi(s) = \varphi(\hat{s})$.

Consider now a signal profile s on the boundary, i.e. $s \in S \backslash int(S)$, that is associated with strict preference $\pi(s)$. Choose any i for whom $s^i \in S^i \backslash int(S^i)$.

Claim 4. *There exists a signal profile $\tilde{s}$ such that $\tilde{s}^i \in int(S^i)$, $\tilde{s}^{-i} = s^{-i}$, and $\pi(s) = \pi(\tilde{s})$.*

Proof. Let $\hat{s}$ be a signal such that $\hat{s}^i \in int(S^i)$ and $\hat{s}^{-i} = s^{-i}$. Because S^i is a convex set, the agents' utility functions are continuous, and $\pi(s)$ is a strict preference profile, there exists $\lambda \in (0,1)$ such that $\tilde{s} := \lambda \hat{s} + (1-\lambda)s$ is in S and $\pi(\tilde{s}) = \pi(s)$. Note that $\tilde{s}^{-i} = s^{-i}$ by definition of $\tilde{s}$. To show that $\tilde{s}^i \in int(S^i)$, first note that there exists $\varepsilon > 0$ such that, for any $\bar{s}^i \in \mathbb{R}^{m^i}$, $\|\bar{s}^i - \hat{s}^i\| < \varepsilon \Rightarrow \bar{s}^i \in S^i$ because $\hat{s}^i \in int(S^i)$ by assumption. This fact and convexity of S^i imply that there exists $\varepsilon' > 0$ such that, for any $\bar{s}^i \in \mathbb{R}^{m^i}$, $\|\bar{s}^i - \tilde{s}^i\| < \varepsilon' \Rightarrow \bar{s}^i \in S^i$.[35] This means that $\tilde{s}^i \in int(S^i)$, completing the proof. □

Let $\tilde{s}$ be a signal profile $\tilde{s}$ such that $\tilde{s}^i \in int(S^i)$, $\tilde{s}^{-i} = s^{-i}$, and $\pi(s) = \pi(\tilde{s})$: such a signal $\tilde{s}$ exists by Claim 4. Then, a proof similar to that in Claim 1 above can be used to show $\varphi(s) = \varphi(\tilde{s})$. Repeating this argument for each i whose signal s^i is on the boundary, we can establish that $\varphi(s) = \varphi(\tilde{s}) = \cdots = \varphi(\hat{s})$ for some $\hat{s} \in int(S)$, which completes the proof. □

Proof of Theorem 3. We only need to check that condition (2) or (3) holds if and only if there exist $p, p' \in [0,1]$ such that each agent with type $\bar{s}^i$ is indifferent between reporting

[35] The claim holds for $\varepsilon' = \lambda\varepsilon$ for instance. To see this, let $\bar{s}^i$ be a point in $\mathbb{R}^{m^i}$ such that $\|\bar{s}^i - \tilde{s}^i\| < \varepsilon'$. Let $\check{s}^i \in \mathbb{R}^{m^i}$ be defined as $\check{s}^i := \frac{\bar{s}^i - (1-\lambda)s^i}{\lambda}$. Because $\hat{s}^i = \frac{\tilde{s}^i - (1-\lambda)s^i}{\lambda}$ by definition of $\tilde{s}^i$, it follows that $\|\check{s}^i - \hat{s}^i\| = \|\bar{s}^i - \tilde{s}^i\|/\lambda < \frac{\varepsilon'}{\lambda} = \varepsilon$, implying that $\check{s}^i \in S^i$. Because $\bar{s}^i$ can be expressed as a convex combination $\bar{s}^i = \lambda \check{s}^i + (1-\lambda)s^i$, this and convexity of S^i imply $\bar{s}^i \in S^i$.

some $s^i > \bar{s}^i$ and some $\tilde{s}^i < \bar{s}^i$. Given this and the fact that u^i is increasing in s^i, each agent i with $s^i > (<)\bar{s}^i$ would prefer reporting truthfully to any $\tilde{s}^i < (>)\bar{s}^i$ so φ^* is Bayesian incentive compatible. Now the indifference condition for $\bar{s}^i$ requires: for agent 1,

$$p' \int_0^{\bar{s}^2} u^1(\bar{s}^1, s^2) dF^2(s^2) = \int_0^{\bar{s}^2} u^1(\bar{s}^1, s^2) dF^2(s^2) + p \int_{\bar{s}^2}^1 u^1(\bar{s}^1, s^2) dF^2(s^2)$$

or

$$(1-p') \int_0^{\bar{s}^2} u^1(\bar{s}^1, s^2) dF^2(s^2) + p \int_{\bar{s}^2}^1 u^1(\bar{s}^1, s^2) dF^2(s^2) = 0 \tag{12}$$

and for agent 2,

$$(1-p') \int_0^{\bar{s}^1} u^2(s^1, \bar{s}^2) dF^1(s^1) = \int_0^{\bar{s}^1} u^2(s^1, \bar{s}^2) dF^1(s^1) + (1-p) \int_{\bar{s}^1}^1 u^2(s^1, \bar{s}^2) dF^1(s^1)$$

or

$$p' \int_0^{\bar{s}^1} u^2(s^1, \bar{s}^2) dF^1(s^1) + (1-p) \int_{\bar{s}^1}^1 u^2(s^1, \bar{s}^2) dF^1(s^1) = 0. \tag{13}$$

Define

$$E_-^i := \int_0^{\bar{s}^{-i}} u^i(\bar{s}^i, s^{-i}) dF^{-i}(s^{-i}) \text{ and } E_+^i := \int_{\bar{s}^{-i}}^1 u^i(\bar{s}^i, s^{-i}) dF^{-i}(s^{-i})$$

and note that $E_-^i < 0$ and $E_+^i > 0$. Then, (12) and (13) can be rewritten as

$$(1-p')E_-^1 + pE_+^1 = 0 \tag{14}$$

$$p'E_-^2 + (1-p)E_+^2 = 0. \tag{15}$$

To check the existence of $(p, p') \in [0, 1]$ that solves these equations, in Figure 4 below, we draw solid lines passing through $(p, p') = (0, 1)$ for (14) and dashed lines passing through $(p, p') = (1, 0)$ for (15).

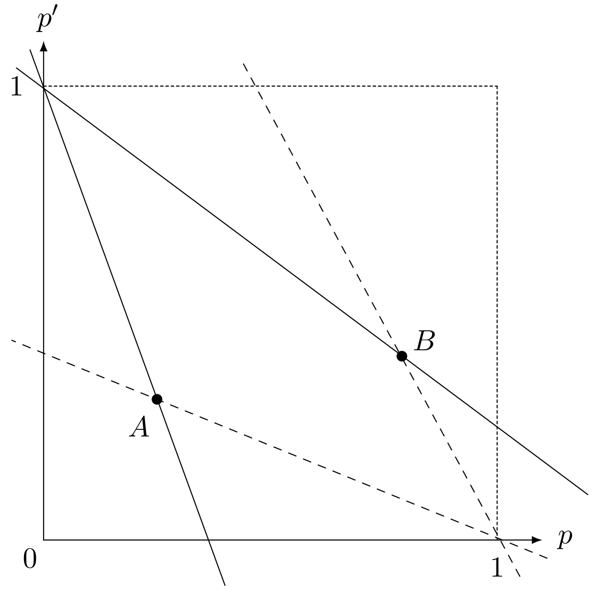

Figure 4

So the linear system (14) and (15) will have a solution if and only if two lines intersect at a point like either A or B. For A to be an intersection, the slope of (14) must be weakly smaller than -1 while the slope of (15) must be weakly greater than -1, i.e.

$$\frac{E_+^1}{E_-^1} \leq -1 \text{ and } \frac{E_+^2}{E_-^2} \geq -1$$

or

$$E_-^1 + E_+^1 \geq 0 \geq E_-^2 + E_+^2,$$

which is equivalent to (2). Similarly, B being an intersection is equivalent to (3). $\square$

Proof of Theorem 4. It suffices to show that any Pareto efficient and ex-post monotonic mechanism must take the same form as φ^* almost everywhere.

Let $S_{NE} := [\bar{s}^1, 1] \times [\bar{s}^2, 1] \subset [0,1]^2$, i.e. the northeast square in the right panel of Figure 2. Let S_{NW}, S_{SE}, and S_{SW} be defined similarly. Note that $S_{ab} \subset S_{SE}$ and $S_{ba} \subset S_{NW}$. Consider any Pareto efficient and ex-post monotonic assignment φ. By the Pareto efficiency, we must have $\varphi_a^1(s) = 1 = 1 - \varphi_a^1(\hat{s})$ for all $s \in S_{ab}, \hat{s} \in S_{ba}$.

Now consider any signal profile $s \in S_{NW} \cap S_{aa}$. We can find another profile $\hat{s} \in S^{ba}$ with $\hat{s}^1 = s^1$ and $\hat{s}^2 \leq s^2$. The ex-post monotonicity and $\varphi_a^2(\hat{s}) = 1$ imply $\varphi_a^2(s) = 1$, and this implies $\varphi_a^1(s) = 0$. Consider alternatively any profile $s \in S_{NW} \cap S_{bb}$. We can find another profile $\hat{s} \in S_{ba}$ with $\hat{s}^1 \geq s^1$ and $\hat{s}^2 = s^2$. The ex-post monotonicity and $\varphi_a^1(\hat{s}) = 0$ imply $\varphi_a^1(s) = 0$. To sum, we must have $\varphi_a^1(s) = 1 - \varphi_a^2(s) = 0$ for all $s \in S_{NW}$. In a similar fashion, it can be shown that $\varphi_a^1(s) = \varphi_a^2(s) = 1$ for all $s \in S_{SE}$.

Let us now turn to the region S_{NE}. We need to show that $\varphi_a^1(s)$ is constant almost everywhere in S_{NE}. Consider any two signals $s^1, \hat{s}^1 \in [\bar{s}^1, 1]$ for agent 1 with $\hat{s}^1 > s^1$. Note that by the above argument

$$\varphi_a^1(s^1, s^2) = \varphi_a^1(\hat{s}^1, s^2) = 1, \forall (s^1, s^2), (\hat{s}^1, s^2) \in S_{SE}. \tag{16}$$

Also, by the ex-post monotonicity,

$$\varphi_a^1(s^1, s^2) \leq \varphi_a^1(\hat{s}^1, s^2), \forall (s^1, s^2), (\hat{s}^1, s^2) \in S_{NE}. \tag{17}$$

Now the difference in expected payoffs of s^1 when reporting s^1 and when reporting $\hat{s}^1$ can be written as

$$\begin{aligned} &\int_0^{\bar{s}^2} (\varphi_a^1(s^1, s^2) - \varphi_a^1(\hat{s}^1, s^2)) u^1(s) dF^2(s^2) + \int_{\bar{s}^2}^1 (\varphi_a^1(s^1, s^2) - \varphi_a^1(\hat{s}^1, s^2)) u^1(s) dF^2(s^2) \\ = &\int_{\bar{s}^2}^1 (\varphi_a^1(s^1, s^2) - \varphi_a^1(\hat{s}^1, s^2)) u^1(s) dF^2(s^2), \end{aligned}$$

where the equality holds since the first integral is equal to zero due to (16). Note that $u^1(s) > 0 \; \forall s \in S_{NE}$. Thus, if (17) holds as a strict inequality for a positive measure of s_2's in S_{NE}, then the payoff difference above would be strictly negative so agent 1 with signal s^1 would be better off reporting $\hat{s}^1$ rather than s^1. So we must have

$$\varphi_a^1(s^1, s^2) = \varphi_a^1(\hat{s}^1, s^2), \text{ for almost all } (s^1, s^2), (\hat{s}^1, s^2) \in S_{NE}. \tag{18}$$

A similar argument can be used to show

$$\varphi_a^2(s^1, s^2) = \varphi_a^2(s^1, \hat{s}^2), \text{ for almost all } (s^1, s^2), (s^1, \hat{s}^2) \in S_{NE}. \tag{19}$$

Combining (18) and (19), we obtain for almost all signal profiles $(s^1, s^2), (\hat{s}^1, \hat{s}^2) \in S_{NE}$,

$$\varphi_a^1(s^1, s^2) = \varphi_a^1(\hat{s}^1, s^2) = 1 - \varphi_a^2(\hat{s}^1, s^2) = 1 - \varphi_a^2(\hat{s}^1, \hat{s}^2) = \varphi_a^1(\hat{s}^1, \hat{s}^2).$$

A symmetric argument can be used to show $\varphi_a^1(s)$ is constant almost everywhere in S_{SW}. □

Proof of Proposition 1. First we write $v^i_{o_k}(s) = \alpha_k w^i(s) + \beta_k$, where $w^i(s) := \gamma s^i + (1-\gamma)\frac{\sum_{j\neq i} s^j}{n-1}$. Fix any $s \in S_\pi$ and $\hat{s} \in S_{\hat{\pi}}$. We show that there exists a step-wise path of the desired form between the two points. We show these in two steps.

Claim 5. *Fix any $s \in S_\pi$ and $\hat{s} \in S_{\hat{\pi}}$ for strict preference profiles π and $\hat{\pi}$. There exists a continuous path $\sigma : [0,1] \to S$ with $\sigma(0) = s$ and $\sigma(1) = \hat{s}$ such that $\sigma(t) \in S_{\pi'}$ for a strict preference profile π' for all $t \in [0,1]$, except possibly for finite values, $\{t_1, ..., t_K\}$, and for each value $t \in \{t_1, ..., t_K\}$, $\sigma(t) \in S_{\pi''}$ for π'' that is strict for all agents except for one. (In words, there exists a continuous path connecting s and $\hat{s}$ that crosses an indifference curve of at most one agent at a time and only finitely many times.)*

Proof. For each agent i, there exists only a finite number of w^i's between $w^i(s)$ and $w^i(\hat{s})$ such that $\alpha_k w^i + \beta_k = \alpha_l w^i + \beta_l$ for some $k \neq l$. This means that for any open neighborhood W containing $w(\hat{s}) := (w^1(\hat{s}), ..., w^n(\hat{s}))$, the set

$$W' := \Bigg\{ w \in W \,\Bigg|\, \exists i, j \in N, i \neq j, \exists k, l, k', l' \in O, k \neq l, k' \neq l', \text{ and } \exists t \in [0,1] \text{ such that } (\alpha_k - \alpha_l)(tw^i + (1-t)w^i(s)) = \beta_l - \beta_k \text{ and } (\alpha_{k'} - \alpha_{l'})(tw^j + (1-t)w^j(s)) = \beta_{l'} - \beta_{k'} \Bigg\}$$

is a lower-dimensional subset of W.[36] In particular, $W \setminus W'$ is nonempty.

Observe next that the gradient matrix $\nabla w(s) := [\nabla_{s^j} w^i]$ has $\gamma > 1/2$ on the diagonal entries and $\frac{1-\gamma}{n-1}$ on the off diagonal entries, so has a full rank. Hence, for an open neighborhood $U \subset S_{\hat{\pi}}$ containing $\hat{s}$, there exists an open neighborhood W containing $w(\hat{s}) := (w^1(\hat{s}), ..., w^n(\hat{s}))$ such that for each $w \in W$ there exists $\tilde{s} \in U$ with $w = w(\tilde{s}) = (w^1(\tilde{s}), ..., w^n(\tilde{s}))$. In particular, one can choose $\tilde{s}$ with $w(\tilde{s}) = w$ for some $w \in W \setminus W'$. The path $\tilde{\sigma}(t) = t\tilde{s} + (1-t)s$ then crosses at most one agent's indifference surface at a given t, and there can be only a finite number of such t's in $[0,1]$. By construction, the path $\hat{\sigma}(t) = t\hat{s} + (1-t)\tilde{s}$ stays inside $S_{\hat{\pi}}$ since the latter set is convex (and hence crosses no agent's indifference surface). Finally, a path $\sigma(t) := \tilde{\sigma}(2t)$ for $t \in [0, \frac{1}{2}]$ and $\sigma(t) := \hat{\sigma}(2t-1)$ for $t \in (\frac{1}{2}, 1]$ satisfies the requirement. □

[36]The fact that W' is a lower-dimensional subset of W can be seen because

$$W' = \bigcup_{i\neq j, k\neq l, k'\neq l'} \left\{ w \in W \,\middle|\, \frac{(\beta_l - \beta_k) - (\alpha_k - \alpha_l)w^i(s)}{(\alpha_k - \alpha_l)(w^i - w^i(s))} = \frac{(\beta_{l'} - \beta_{k'}) - (\alpha_{k'} - \alpha_{l'})w^j(s)}{(\alpha_{k'} - \alpha_{l'})(w^j - w^j(s))} \in [0,1] \right\}$$

and the latter set is clearly lower-dimensional.

Claim 6. *There exists a step-wise path of the required form connecting s and $\hat{s}$.*

Proof. By Claim 5, there exists a continuous path $\sigma(t)$ with $\sigma(0) = s$ and $\sigma(1) = \hat{s}$ such that σ crosses at most one agent's indifference surface at a given t, only for finitely many t's: $0 < t_1 < ... < t_{K-1} < 1$ for some positive integer K. Let $t_0 \equiv 0$ and $t_K \equiv 1$. Let agent $i_k \in N$ be indifferent over at least a pair of objects at t_k, $0 < k < K$, and let π_k be such that $\sigma(t) \in S_{\pi_k}$ for all $t \in (t_{k-1}, t_k)$. Then, since the specified utility function satisfies the Interdependence assumption, for such k there exists $j_k \neq i_k$ such that $s_{k-} \equiv \sigma(t_k) - \varepsilon e^{j_k} \in S_{\pi_k}$ and $s_{k+} \equiv \sigma(t_k) + \varepsilon e^{j_k} \in S_{\pi_{k+1}}$ for a (positive or negative) real number ε with a sufficiently small absolute value, where e^j is a vector whose component corresponding to j equals one and all other components equal zero. For ε with any sufficiently small absolute value, any signal on the line segment between s_{k-} and s_{k+} gives rise to the same strict preference for all agents, except for agent i_k whose preferences change from $\pi_k^{i_k}$ to $\pi_{k+1}^{i_k}$ as one moves from s_{k-} to s_{k+} along that line segment. Since a set S_{π_k} is an open connected set, by Lemma 1, there exists a step-wise path $\bar{\sigma}_k$ connecting $s_{(k-1)+}$ and s_{k-}, which varies in one agent's signal on each segment and lies within S_{π_k} (where $s_{0+} \equiv s$ and $s_{K-} \equiv \hat{s}$). Piecing together $\bar{\sigma}_k$ with the line segment $\overline{s_{k-}s_{k+}}$, for each $k = 1, ..., K-1$ and finally connecting with $\bar{\sigma}_K$, we construct a step-wise path of the required form. □

The above claims prove Proposition 1. □

Appendix B: Relationships across Different Conditions

The Rich Domain assumption does not imply the Rich Domain* assumption. As explained in the main text of the paper, the canonical one-dimensional signal model of Example 1 for any $n \geq 3$ is a counterexample.

The Rich Domain* assumption does not imply the Rich Domain assumption. To see this point, consider a model with three agents $1, 2, 3$ and three objects. For each i, her ordinal preference depends only on signal s^{i+1} (where we use the convention that $i + 1 = 1$ for $i = 3$). Moreover, assume that for each i and her ordinal preference π^i, there exists $s^{i+1} \in int(S^{i+1})$ such that the ordinal preference is π^i when the signal of agent $(i + 1)$ is s^{i+1}. Then the Rich Domain* assumption is satisfied by definition. On the other hand, for any $i, j \in N$, and s^{-ij}, the ordinal preference of either i or j is constant across all s^i and s^j by construction, implying that the Rich Domain assumption is violated.

FK: In the definition of Connectedness, we say "Consider i, j, ..." but there could be more than one such choice. I think that we should clarify that "Consider **any** ..." or "There exists a signal profile (and other stuffs) that satisfy the Rich Domain..." (the latter seems enough for our main result, but in that case we should rewrite the proof a bit). The Connectedness assumption does not imply the Connectedness* assumption. To see this, consider a model with three agents $1, 2, 3$ and three objects a, b, c, with the signal space of each agent being $[0, 1]$. The utility function of agent $i \in \{1, 2\}$ is given by $v_a^i(s) = 0$, $v_b^i(s) = s^i - \frac{1}{2}$, and $v_c^i(s) = -1$ for all s. For agent $j = 3$, let $v_a^j(s) = -1$, $v_b^j(s) = (s^3 - \frac{1}{2})^2 - \frac{1}{8}$, and $v_c^j(s) = 0$ for all s. Then clearly the Connectedness assumption holds with a, b, $i = 1, j = 2$ and any $s^{-ij} = s^3$ in the definition of the condition, while Connectedness* is violated since the subset of the signal space at which agent 3 prefers b to c (and c to a) is not connected.

The Connectedness* assumption does not imply the Connectedness assumption. To see this, consider a model with three agents $1, 2, 3$ and three objects a, b, c, with the signal space of each agent $[0, 1]$. For agent $i \in \{1, 2\}$ is $v_a^i(s) = 0$, $v_b^i(s) = (s^i - \frac{1}{2})^2 - \frac{1}{8} + s^3$, and $v_c^i(s) = -1$ for all s. The preference of agent 3 is constant across signals. Then Connectedness* holds, but Connectedness is violated because the Rich Domain assumption can only be satisfied with objects a and b, agents 1 and 2, and $s^3 < 1/8$, in which case, however, the set of agents 1 and 2's signals for which they prefer a to b and b to c is not connected.