GENERALIZED REDUCED-FORM AUCTIONS: A NETWORK-FLOW APPROACH

YEON-KOO CHE, JINWOO KIM, AND KONRAD MIERENDORFF

ABSTRACT. We develop a network-flow approach for characterizing interim-allocation rules that can be implemented by ex post allocations. The network method can be used to characterize feasible interim allocations in general multi-unit auctions where agents face hierarchical capacity constraints. We apply the method to solve for an optimal multi-object auction mechanism when bidders are constrained in their capacities and budgets.

KEYWORDS: Reduced-form auctions, network-flow approach, Gale's demand theorem, hierarchy of capacity constraints.
JEL-CODE: D44.

1. INTRODUCTION

In the classical auction design problem, a bidder's (Bayesian) incentive constraint is used to express his payoff (and thus his payment) in terms of interim allocations—his expected winning probabilities given his types. This allows one to eliminate the payment rule in the seller's objective function, thereby reducing the problem to that of optimizing solely over allocation rules. This celebrated method works since in the last step the optimal allocation can be searched point-wise, for each type profile (Myerson, 1981).

There are situations, however, in which the allocation rule cannot be optimized point-wise for each type profile. For instance, agents may face constraints in their payments for a variety of reasons.[1] Given the envelope condition, such payment constraints can be readily checked for an interim allocation, but not for an ex post allocation. A similar situation is

---

Date: September 27, 2011.
Che: Department of Economics, Columbia University and YERI, Yonsei University, yeonkooche@gmail.com.
Kim: Department of Economics, Yonsei University, jikim72@gmail.com.
Mierendorff (corresponding author): Chair of Organizational Design, Department of Economics, University of Zürich, konrad.mierendorff@econ.uzh.ch. Che and Kim acknowledge the support by WCU program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R32-2008-000-10056-0). Mierendorff gratefully acknowledges financial support from the European Research Council (ERC Advanced Investigator Grant, ESEI-249433). Part of this research was conducted while Che visited the University of Zürich and Mierendorff visited Yonsei University; they are grateful for the hospitality of the hosting institutions.

[1]For instance, the agents may be financially constrained (Che and Gale, 1998, 2000; Laffont and Robert, 1996; Maskin, 2000; Pai and Vohra, 2008). In the context of collusion agreement, members of a cartel may refrain from using monetary transfers, for fear of detection (McAfee and McMillan, 1992; Che et al., 2010). Or monetary transfers may be simply unavailable for other reasons (Miralles, 2008; Che et al., 2011).

encountered if agents have type-contingent outside options (Mierendorff, 2009). Again, this constraint can be checked for agents' interim allocations (via the envelope expression), but not for ex post allocations. Consequently, these situations require maximizing directly over interim allocation rules to find the optimal mechanism.[2]

For this approach to work, however, one must have a handle on the following issue: *what interim allocation rules are implementable (or so called reduced form auctions), in the sense that there exists an ex post allocation rule generating the desired interim winning probabilities?* Proving a conjecture by Matthews (1984), Border (1991, 2007) characterized implementable interim allocations for the one-unit auctions case.[3,4]

The current note revisits implementability of reduced-form auctions using a network-flow approach: We map the problem of existence of an ex-post allocation rule that implements given interim winning probabilities, to the problem of the existence of a feasible flow on an appropriately defined network. Gale's demand theorem (Gale, 1957), which is based on the Max-Flow-Min-Cut theorem (Ford and Fulkerson, 1956), provides a necessary and sufficient condition for the existence of a feasible flow. We find that this condition boils down to Border's characterization for the single-unit auction setting. We thus provide a network-flow interpretation of the well-known result, and in the process make the insight behind the implementability condition more transparent.[5]

More importantly, our network-flow approach "unweds" the implementability of reduced-form allocations from the single-unit auction setting, thus paving way to asking the implementability question in a much broader environment. Indeed, we characterize implementable interim allocations in a general multi-unit environment in which subsets of agents may be subject to capacity constraints, as long as these subsets are nested.[6] For a single agent, such a capacity constraint may arise from his limited ability to utilize the units he obtains. For instance, firms can profitably utilize at most so many units (e.g., spectrum licenses). Constraints on groups of agents may arise from the seller's (e.g., the government's) desire

[2]Maskin and Riley (1984) were the first to take such an approach. See also Armstrong (2000), Asker and Cantillon (2010), and Parlane (2001). In different contexts, Border's results have also been used by Brusco and Lopomo (2002), Manelli and Vincent (2010), and Hörner and Samuelson (2011).

[3]For the case of asymmetric agents, Mierendorff (2011) and Che et al. (2010) offer a tighter characterization than Border (2007).

[4]In the case of two buyers, the problem of implementing a given reduced form corresponds to the problem of finding a two-dimensional distribution with given marginals. The classic papers on this problem are Lorentz (1949), Kellerer (1961) and Strassen (1965). The characterization of asymmetric reduced forms for the two-buyer case is a direct consequence of their results. These methods, however, are not applicable in the case of more than two buyers. We thank Benny Moldovanu for pointing us to this literature (see also Gershkov et al., 2011). Recently, Hart and Reny (2011) have shown that for symmetric buyers, Border's characterization is equivalent to a second-order stochastic dominance condition.

[5]Gershkov et al. (2011) point out a connection between implementability and a majorization condition found in Gale (1957) (see also Ryser, 1957). Gale's construction, however, only works for two buyers. Also, his condition differs from Border's and seems less tractable.

[6]That is, the subsets form a hierarchy: any two subsets are either disjoint, or one is a subset of the other.

to nurture minority participation or to preserve a competitive (post-assignment) industry. For instance, the government may want to limit the number of units allocated to large or incumbent firms, "setting aside" some units for small firms or new entrants.

We begin in Section 2 with an example that illustrates the network-flow approach. In Section 3, we present the general model with capacity constraints on hierarchical sets of agents and give a characterization of implementable reduced forms. These generalized implementability conditions are still very rich. In the spirit of Border (1991) and Mierendorff (2011), we reduce the conditions to those defined over upper-contour sets in Section 4. We also show how the conditions are further reduced in case of symmetry. In Section 5, we provide an application where agents face both capacity and budget constraints, which further illustrates the relevance of the reduced-form auction approach as well as the use of our characterization.

2. Examples

2.1. **The Network Approach.** In this section, we illustrate by a simple example how the problem of characterizing reduced form allocation rules can be mapped to a network flow problem. Suppose that a seller has $C = 3$ units of a good and that there are two potential buyers $i = 1, 2$. The type-space of each agent is binary, $\Theta_i = \{\underline{\theta}_i, \bar{\theta}_i\}$. The probability that agent i has type θ_i is denoted $p_i(\theta_i)$.[7] An **(ex-post) allocation rule** is a pair of functions $Q_i : \Theta \to [0, C]$, $i = 1, 2$, where $\Theta = \Theta_1 \times \Theta_2$, such that

$$\forall \theta \in \Theta : \qquad Q_1(\theta) + Q_2(\theta) \leq C. \tag{2.1}$$

$Q_i(\theta)$ describes the "fractional" assignment of units to agent i for a given type profile θ.

A given ex post allocation rule (Q_1, Q_2) induces an **interim allocation rule**, $q_i : \Theta_i \to [0, C]$, $i = 1, 2$, representing i's expected assignment given his type: for each i and θ_i,

$$q_i(\theta_i) \quad = \quad Q_i(\theta_i, \underline{\theta}_{-i}) p_{-i}(\underline{\theta}_{-i}) + Q_i(\theta_i, \bar{\theta}_{-i}) p_{-i}(\bar{\theta}_{-i}). \tag{2.2}$$

Conversely, one could begin with an arbitrary interim allocation rule and ask whether it can be implemented by an ex post allocation rule. As motivated in the introduction, such an approach is necessary in certain situations. Formally, an interim allocation rule $q_i : \Theta_i \to [0, C]$, $i = 1, 2$, is **implementable** if it is the **reduced form** of an ex-post allocation rule, i.e., if there exist Q_i satisfying (2.1) and (2.2).

For the case of one good, Border (2007) shows that (q_1, q_2) is implementable if and only if for each $T_1 \subset \Theta_1$ and $T_2 \subset \Theta_2$, the probability that an agent i with type $\theta_i \in T_i$ wins is less than or equal to the probability that there is such an agent. Formally, for all $T_1 \subset \Theta_1$ and

[7]For the example, we assume independence but this is not necessary to apply the network approach.

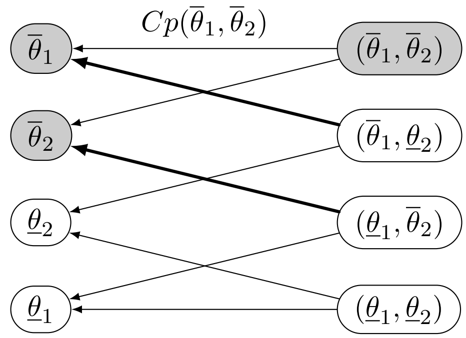

FIGURE 2.1. Illustration of the network (S = the set of shaded nodes).

$T_2 \subset \Theta_2$,

$$\sum_{i=1,2} \sum_{\theta_i \in T_i} q_i(\theta_i) p_i(\theta_i) \le \sum_{\theta \in [T_1 \times \Theta_2] \cup [\Theta_1 \times T_2]} p(\theta) = 1 - \sum_{\theta \in T_1^c \times T_2^c} p(\theta),$$

where $p(\theta) = p_1(\theta_1)p_2(\theta_2)$, and $T_i^c = \Theta_i \setminus T_i$ is the complement of T_i. For C units, the condition becomes

$$\sum_{i=1,2} \sum_{\theta_i \in T_i} q_i(\theta_i) p_i(\theta_i) \le C \sum_{\theta \in [T_1 \times \Theta_2] \cup [\Theta_1 \times T_2]} p(\theta) = C \left(1 - \sum_{\theta \in T_1^c \times T_2^c} p(\theta) \right). \quad (2.3)$$

The left-hand side of this condition can be interpreted as the ex ante expected quantity that is allocated to agents i with types $\theta_i \in T_i$. The right-hand side can be interpreted as the ex-ante expected supply that is available in states where at least one agent i has a type $\theta_i \in T_i$.

A network flow approach can be employed to show that this condition is necessary and sufficient for implementability. To this end, we define a directed network as follows:

The network consists of nodes N and directed edges E defined over pairs of nodes. Figure 2.1 illustrates the network for our example. There are several types of nodes in N. "Demand nodes" $D = \Theta_1 \cup \Theta_2 (= \{\underline{\theta}_1, \bar{\theta}_1, \underline{\theta}_2, \bar{\theta}_2\})$ are given by possible agent-type pairs θ_i.[8] Demand nodes are arranged on the left side of Figure 2.1. With each demand node we associate a **demand** for the good given by $d(\theta_i) = p_i(\theta_i)q_i(\theta_i)$. Intuitively, interim expected quantities (q_1, q_2) specify the expected units required by each agent-type pair. Hence, the demand $d_i(\theta_i)$ represents the expected units that the agent type θ_i must receive so as to implement the desired interim allocation rule. On the right side, there are "supply nodes" which correspond to the set of type profiles, or more aptly the set of "states," Θ. With each supply node $\theta \in \Theta$, we associate a supply (negative demand) of the good given by $d(\theta) = -p(\theta)C$. Intuitively, there is a total supply of C, and with probability $p(\theta)$, this supply is available in state θ.

An interim allocation rule is implementable if the supply available in each state can be allocated to satisfy the demand given by that interim allocation rule. We shall check the

[8]For convenience of notation, we identify θ_i with the agent-type pair (i, θ_i).

latter by asking if a sufficient flow can be sent from the supply nodes to the demand nodes to fulfill the demand required by the interim quantities. Of course, the supply available at a given state θ cannot be used to fulfill the demand at all demand nodes. For instance, in state $(\underline{\theta}_1, \underline{\theta}_2)$, units can be allocated to an agent $i = 1, 2$ with type $\underline{\theta}_i$ but not to an agent i with type $\bar{\theta}_i$. Directed edges $E \subset N \times N$ describe to which demand nodes the supply at different states can be distributed. For instance, there are two edges from $(\underline{\theta}_1, \underline{\theta}_2)$ to the demand nodes $\underline{\theta}_1$ and $\underline{\theta}_2$, respectively, but no edges to the demand nodes $\bar{\theta}_1$ and $\bar{\theta}_2$. More generally, supply at $\theta = (\theta_1, \theta_2)$ can only be used to satisfy demand at nodes $\tilde{\theta}_i$ for which $\tilde{\theta}_i = \theta_i$, so there is a directed edge from θ to $\tilde{\theta}_i$ if and only if $\tilde{\theta}_i = \theta_i$. Each directed edge is associated with a capacity $c : E \to \mathbb{R}_+$. In the current example, the entire supply available at θ can be allocated to any demand node θ_i that is connected to θ by a directed edge. Hence, we define the capacity of the edges by $c(\theta, \theta_i) = Cp(\theta)$.

A flow on the network is a function $f : E \to \mathbb{R}_+$, such that $f(e) \leq c(e)$. A flow is called **feasible** given the demand d, if (a) for each supply node $\theta = (\theta_1, \theta_2)$, the sum of the flows on the edges leaving θ is less than or equal to the supply at θ, i.e., $f(\theta, \theta_1) + f(\theta, \theta_2) \leq -d(\theta)$, and (b) for each demand node θ_i, the sum of the flows on the edges entering θ_i must be greater than or equal to the demand at θ_i, i.e., $f((\theta_i, \underline{\theta}_{-i}), \theta_i) + f((\theta_i, \bar{\theta}_{-i}), \theta_i) \geq d(\theta_i)$.

A flow f induces an allocation rule via $Q_i(\theta) = f(\theta, \theta_i)/p(\theta)$. Importantly, if there exists a feasible flow given demand d, then the interim allocation rule associated with that demand is implementable. To see this, fix an interim allocation rule $q_i(\cdot)$ and define demand as above. If there exists a feasible flow given demand d, then we must have $f((\theta_i, \underline{\theta}_{-i}), \theta_i) + f((\theta_i, \bar{\theta}_{-i}), \theta_i) \geq q(\theta_i)p_i(\theta_i)$. Moreover we can reduce the flow on some edges so that the inequality holds with equality for all θ_i. For this reduced flow $\tilde{f}$, define the allocation rule $Q_i(\theta) := \tilde{f}(\theta, \theta_i)/p(\theta)$. Substituting this into the equation, we get

$$Q_i(\theta_i, \underline{\theta}_{-i})p_{-i}(\underline{\theta}_{-i}) + Q_i(\theta_i, \bar{\theta}_{-i})p_{-i}(\bar{\theta}_{-i}) = q_i(\theta_i).$$

Hence, the reduced flow $\tilde{f}$ induces an allocation rule that satisfies (2.2) and thus implements the reduced form q.

Conversely, if we fix an allocation rule Q and define demand by the reduced form of Q, then the flow given by $f(\theta, \theta_i) = Q_i(\theta)p(\theta)$ is feasible given demand d. Therefore, an interim allocation rule $(q_i)_{i=1,2}$ is implementable if and only if there exists a feasible flow for the corresponding network defined above.

Gale's (1957) demand theorem gives conditions for the existence of a feasible flow.

Theorem 1 (Gale's Demand Theorem)**.** *A feasible flow f exists if and only if for each $S \subset N$,*

$$d(S) \leq c(N \setminus S, S), \tag{2.4}$$

where $d(S) := \sum_{x \in S} d(x)$ and $c(S, S') := \sum_{(x,x') \in (S \times S') \cap E} c(x, x')$.

In words, a feasible flow exists if net demand at any subset of nodes S does not exceed the sum of capacities on the edges that enter the nodes in S from the nodes outside S. For example, fix $S = \{\bar{\theta}_1, \bar{\theta}_2, (\bar{\theta}_1, \bar{\theta}_2)\}$. This set consists of the shaded nodes in Figure 2.1. The aggregate demand at these nodes is $d(S) = q_1(\bar{\theta}_1)p_1(\bar{\theta}_1) + q_2(\bar{\theta}_2)p_2(\bar{\theta}_2) - C\,p(\bar{\theta}_1, \bar{\theta}_2)$. The bold-faced edges enter S from outside, and the total capacity of these edges is $c(N \setminus S, S) = C\,(p(\bar{\theta}_1, \underline{\theta}_2) + p(\underline{\theta}_1, \bar{\theta}_2))$. Hence, $d(S) \le c(N \setminus S, S)$ is equivalent to

$$q_1(\bar{\theta}_1)p_1(\bar{\theta}_1) + q_2(\bar{\theta}_2)p_2(\bar{\theta}_2) \le C(p(\bar{\theta}_1, \bar{\theta}_2) + p(\bar{\theta}_1, \underline{\theta}_2) + p(\underline{\theta}_1, \bar{\theta}_2)). \tag{2.5}$$

Note that this corresponds to inequality (2.3) for $T_1 = \{\bar{\theta}_1\}$ and $T_2 = \{\bar{\theta}_2\}$. It turns out that the inequalities generated by adding additional supply nodes to S or by removing $(\bar{\theta}_1, \bar{\theta}_2)$ from S are all implied by (2.5). The inequalities (2.3) for other sets T_i can be derived similarly by considering sets $S \subset N$ that include the demand nodes $\theta_i \in T_1 \cup T_2$. The set of non-redundant inequalities thus generated by Gale's demand theorem is identical to Border's constraints. This proves that (2.3) is necessary and sufficient for the existence of an allocation rule that implements a given interim allocation rule in our example.

2.2. **Introducing additional constraints.** The network approach illustrated above makes transparent the insight associated with Border's conditions. More importantly, it allows us to study reduced-form auctions in a broader environment beyond the simple auction setting. As will be illustrated here and shown more generally in the next section, it can accommodate a fairly general form of capacity constraints on the part of the agents.

To illustrate, we modify the previous example by assuming that, ex post, each agent may get at most two units. An allocation rule therefore has to satisfy the additional constraint that $Q_i(\theta) \le C_i = 2$ for $i = 1, 2$ and all type profiles $\theta \in \Theta$. It turns out that the Border constraints from the previous example are not sufficient for implementability under this additional constraint. To see this, suppose that $p_1(\bar{\theta}_1) = p_2(\bar{\theta}_2) = p > 2/3$ and consider the interim winning probabilities given by $q_1(\bar{\theta}_1) = q_2(\bar{\theta}_2) = \bar{q} = 3 - (3/2)p$ and $q_1(\underline{\theta}_1) = q_2(\underline{\theta}_2) = \underline{q} = (3/2)(1-p)^2$. It is straightforward to check that these interim winning probabilities satisfy the Border constraints. Yet, there is no allocation rule that implements these winning probabilities and satisfies the additional constraint $Q_i(\theta) \le 2$. To see this, note that $Q_i(\bar{\theta}_1, \bar{\theta}_2) \le 3/2$ for at least one i, which follows from $Q_1(\bar{\theta}_1, \bar{\theta}_2) + Q_2(\bar{\theta}_1, \bar{\theta}_2) \le 3$. Using this and $Q_i(\bar{\theta}_i, \underline{\theta}_{-i}) \le 2$, we have

$$\bar{q} = pQ_i(\bar{\theta}_i, \bar{\theta}_{-i}) + (1-p)Q_i(\bar{\theta}_i, \underline{\theta}_{-i}) \le \frac{3}{2}p + 2(1-p) = 2 - \frac{1}{2}p < 3 - \frac{3}{2}p,$$

which is a contradiction.

We need to modify the network defined for the unconstrained case in order to produce the correct condition for implementability. To reflect the additional constraint, we reduce the capacity of each edge (θ, θ_i) to $C_i p(\theta)$.[9] A flow in this network now induces an allocation rule that satisfies the additional capacity constraints on the individual agents. Again, we can invoke Gale's demand theorem to derive a set of inequalities that characterize implementable reduced forms. For example, for $S = \{\bar{\theta}_1, \bar{\theta}_2, (\bar{\theta}_1, \bar{\theta}_2)\}$, the demand at S is $d(S) = q_1(\bar{\theta}_1)p_1(\bar{\theta}_1) + q_2(\bar{\theta}_2)p_2(\bar{\theta}_2) - 3\,p(\bar{\theta}_1, \bar{\theta}_2)$, just as before, but the total capacity of the edges entering S is now $c(N \setminus S, S) = 2\,(p(\bar{\theta}_1, \underline{\theta}_2) + p(\underline{\theta}_1, \bar{\theta}_2))$. The key observation here is that, out of 3 units available at each supply node in $N \setminus S$, say $(\bar{\theta}_1, \underline{\theta}_2)$, at most two units can be distributed to satisfy demand in S, since at most one agent in S (agent 1 in this case) is in a position to receive supply and each agent cannot receive more than two units.

As is easily verified, condition $d(S) \leq c(N \setminus S, S)$ is equivalent to

$$q_1(\bar{\theta}_1)p_1(\bar{\theta}_1) + q_2(\bar{\theta}_2)p_2(\bar{\theta}_2) \leq 3p(\bar{\theta}_1, \bar{\theta}_2) + 2(p(\bar{\theta}_1, \underline{\theta}_2) + p(\underline{\theta}_1, \bar{\theta}_2)). \tag{2.6}$$

Comparing (2.6) to the old constraint (2.5), we observe that the new constraint is more demanding. In fact, (2.6) provides the right condition in this example. This will be seen formally in the next section where we provide the full characterization of implementable interim allocation rules in a general environment.

3. Reduced-Form Auctions with Capacity Constraints

Let $I = \{1, \ldots, |I|\}$ be the set of agents with typical elements $i, j \in I$. For each agent i, there is a finite set of types Θ_i with typical element $\theta_i \in \Theta_i$. As usual, we define $\Theta = \times_{i \in I} \Theta_i$ and $\Theta_{-i} = \times_{j \neq i} \Theta_j$. It will be convenient to identify θ_i with the agent-type pair (i, θ_i). Hence, we define $\Theta^D := \bigcup_{i \in I} \Theta_i$ for all possible agent-type pairs, and $\Theta^D_{-i} := \bigcup_{j \neq i} \Theta_j$ for all agent-type pairs not involving agent i. Note that a typical element of Θ is a type profile $\theta = (\theta_1, \ldots, \theta_{|I|})$, whereas a typical element of Θ^D or Θ^D_{-i} is an agent-type pair θ_j, where $j \in I$ if $\theta_j \in \Theta^D$ and $j \neq i$ if $\theta_j \in \Theta^D_{-i}$. For $\theta \in \Theta$, $p(\theta) \in [0, 1]$ denotes the probability that this type profile, or "state", is realized. The marginal distribution of types of any agent i is denoted by $p_i(\theta_i)$ and the probability of a type profile (θ_i, θ_{-i}) conditional on θ_i is denoted by $p_{-i}(\theta_{-i}|\theta_i) = p(\theta_i, \theta_{-i})/p_i(\theta_i)$.

We formulate capacity constraints as follows. Let $\mathcal{H}$ be a family of subsets of I that includes I and all singleton sets, i.e., $\{i\} \in \mathcal{H}$ for all $i \in I$. For each $G \in \mathcal{H}$, the total number of units that can be allocated to agents $i \in G$ is constrained to be less than C_G, i.e., $\sum_{i \in G} Q_i(\theta) \leq C_G$ for all $\theta \in \Theta$. C_I denotes the total number of units available to the seller. To model the constraints in a network, we assume that the constraint structure is hierarchical: for $G, G' \in \mathcal{H}$ we require that either $G \subset G'$, or $G' \subset G$, or $G \cap G' = \emptyset$.

[9] Note that this approach is quite flexible. For example we could have a different constraint for agent one and agent two by setting different capacities on the edges pointing to demand nodes of agent one (θ, θ_1) and those pointing to demand nodes for agent two (θ, θ_2). In the following section, we will also demonstrate how constraints on the total number of objects that a subset of agents may obtain, can be incorporated in the network approach.

If $G \subsetneqq G'$ and there exists no $G'' \in \mathcal{H}$, such that $G \subsetneqq G'' \subsetneqq G'$, then G' is called the **predecessor** of G. Capacity constraints for individual agents are denoted by C_i. If there is no constraint for some individual agent i, we define C_i to be the capacity constraint of i's predecessor. Without loss of generality, we assume C_G to be less than or equal to $C_{G'}$, where G' is G's predecessor. Finally, let $\mathcal{C} := \{C_G\}_{G \in \mathcal{H}}$.

Several realistic situations are accommodated by the hierarchical capacity constraints. In the example of Section 2.1, we have $I = \{1, 2\}$ and this is the only set for which we have a capacity constraint. Hence, $\mathcal{H} = \{I, \{1\}, \{2\}\}$ and $C_I = 3\,(= C_i)$. In the example of Section 2.2, each agent faces an individual capacity constraint. In this case, the hierarchy of sets $\mathcal{H}$ remains the same but we set $C_i = 2$ and $C_I = 3$.

Another special case of hierarchical capacity constraints arises when the seller wants to limit the number of units accruing to a certain groups of agents such as incumbents or foreign firms. If there are K such groups with members G_k and $G_k \cap G_l = \emptyset$ for $k \neq l$, the corresponding hierarchy of constraints would be given by $\mathcal{H} = \{I, G_1, \ldots, G_K, \{1\}, \ldots, \{|I|\}\}$.

We say an (ex-post) allocation rule $Q : \Theta \to \times_{i \in I}[0, C_i]$ **respects** $(\mathcal{H}, \mathcal{C})$, if for each $G \in \mathcal{H}$, $\sum_{i \in G} Q_i(\theta) \leq C_G$ for all $\theta \in \Theta$. An interim allocation rule $q = (q_1, .., q_n)$, $q_i : \Theta_i \to [0, C_i]$, is *implementable* or a *reduced form* if there exists an allocation rule that respects $(\mathcal{H}, \mathcal{C})$ and satisfies

$$q_i(\theta_i) = \sum_{\theta_{-i} \in \Theta_{-i}} Q_i(\theta_i, \theta_{-i})\, p_{-i}(\theta_{-i}|\theta_i), \forall i \in I. \tag{3.1}$$

For a given constraint structure $(\mathcal{H}, \mathcal{C})$, we define a network (N, E, c). The network is defined slightly differently relative to the previous section to facilitate our proof of the characterization theorem. The node set is now given by $N = \{(G, \theta) \,|\, G \in \mathcal{H}, \theta \in \Theta\} \cup \Theta^D \cup \{\sigma\}$, where each $\theta_i \in \Theta^D$ is called a **demand node** and σ is called **super-source**. Node (G, θ) for $G \in \mathcal{H}$ is called a **capacity node** and we sometimes write (i, θ) instead of $(\{i\}, \theta)$. To simplify notation, for each $\theta \in \Theta$, we introduce capacity node (I, θ) that replaces the supply node θ in the example of Section 2.

Next, the network has the set E of directed edges defined over pairs of nodes. The edges form a tree with the super source σ at its root. Specifically, for each $\theta \in \Theta$, there is a directed edge $(\sigma, (I, \theta)) \in E$ with capacity $c(\sigma, (I, \theta)) = C_I p(\theta)$ that emanates from the super-source and enters the capacity node (I, θ). The edge capacity $C_I p(\theta)$ specifies the maximum ex ante expected units that are available at state θ for all agents.[10] Next, for each capacity node (G, θ), $G \neq I$, there is a directed edge $((G', \theta), (G, \theta)) \in E$, if G' is the predecessor of G, with capacity $c((G', \theta), (G, \theta)) = C_G p(\theta)$. The capacity specifies the maximum ex

[10]In other words, the edges do not originate from supply nodes Θ but rather from the super source, unlike the illustrative example in Section 2. For this reason, $C_I p(\theta)$ is now encoded as the capacity of the edge from σ to (I, θ), rather than a supply (or a negative demand) at the supply node θ.

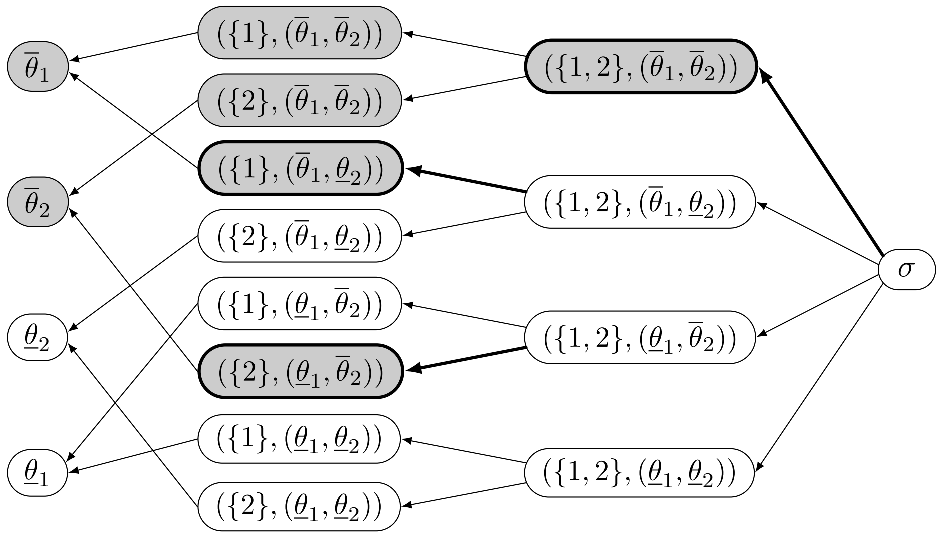

FIGURE 3.1. Bold nodes constitute roots and shaded nodes constitute $S(T)$, where $T_i = \{\overline{\theta}_i\}, \forall i$ while capacity constraints are $C_i = 2, \forall i$ and $C_{\{1,2\}} = 3$.

ante units that group G can receive in state θ (in expectation). Finally, for each demand node θ_i, and each $\theta_{-i} \in \Theta_{-i}$, there is a directed edge $((i,(\theta_i,\theta_{-i})),\theta_i) \in E$ with capacity $c((i,(\theta_i,\theta_{-i})),\theta_i) = \infty$. The unlimited capacity assumption is for analytical convenience, and it does not preclude an individual capacity limit, which is encoded as the capacity of the edge entering capacity node (i,θ). As before, $((i,(\theta_i',\theta_{-i})),\theta_i) \notin E$ if $\theta_i' \neq \theta_i$.

For a given interim allocation rule $q = (q_i : \Theta_i \to [0, C_i])_{i \in I}$, we define demand as $d(\theta_i) = p_i(\theta_i)q_i(\theta_i)$ for demand nodes, $d(G,\theta) = 0$ for capacity nodes, and $d(\sigma) = -\infty$.[11] Figure 3.1 illustrates this network for the example in Section 2.2. Note that in addition to the super-source, we have also introduced an additional layer of capacity nodes preceding the demand nodes. This simplifies the notation and the exposition of the proofs.

A flow on (N, E, c) is a function $f : E \to \mathbb{R}_+$ such that for all $e \in E$, $f(e) \leq c(e)$, and for each capacity node $(G,\theta) \in N$, the flow entering (G,θ) equals the flow leaving it. Formally, the following flow conservation condition must hold for any capacity node n:

$$\sum_{n'':(n,n'') \in E} f(n,n'') = f(n',n), \text{ where } (n',n) \in E.$$

A flow is **feasible** for given demand d if for all demand nodes $\theta_i \in \Theta^D$,[12]

$$\sum_{\theta_{-i} \in \Theta_{-i}} f((i,(\theta_i,\theta_{-i})),\theta_i) \geq d(\theta_i).$$

[11]The infinite supply at σ is assumed for analytical ease in the proof and has no consequence otherwise; the capacities of the edges emanating from σ clearly constrain the number of units that can be allocated to the agents.

[12]Since we define the supply of the super-source as $d(\sigma) = -\infty$, the aggregate flow leaving σ cannot exceed the available supply.

The following lemma operationalizes our network flow framework.

Lemma 1. *An interim allocation is implementable if and only if there exists a feasible flow for the demand given by that interim allocation rule.*

By Gale's demand theorem, a feasible flow exists if and only if the network satisfies (2.4) for each $S \subset N$. We exploit this result to characterize implementable reduced-form auctions. Our goal is to derive a set of conditions for implementable reduced-forms, *one* for *each* profile of type sets $T = (T_1, ..., T_{|I|}), T_i \in \Theta_i$, much in the spirit of Border (1991, 2007). To this end, it is natural to consider (2.4) for sets S that contain the demand nodes corresponding to these types sets, i.e., $T^D = \cup_i T_i \subset \Theta^D$. There are many sets S that satisfy $S \cap \Theta^D = T^D$. For each T^D, we identify the set S that gives rise to the tightest condition. The associated Gale conditions then yield the desired characterization.

To do so, we need some more definitions. In particular, for a given profile of type sets $T = (T_1, ..., T_{|I|}), T_i \subset \Theta_i$, we need to describe which agents can receive the good in each state $\theta \in \Theta$, and what is the most they can collectively receive at that state, if their types are in T. To begin, for any $I' \subset I$, let us call a family of sets $H \subset \mathcal{H}$ a **cover** of I' if $I' \subset \bigcup_{G \in H} G$. The set of all covers of I' is denoted by $\mathcal{P}(I') := \left\{ H \subset \mathcal{H} \,\middle|\, I' \subset \bigcup_{G \in H} G \right\}$. If we add up the capacities C_G for all sets $G \in H$ in a cover of I', we obtain an upper bound for the total number of units that the agents in I' may obtain.

Again fix the profile of type sets $T = (T_1, \ldots, T_{|I|})$. For any state $\theta \in \Theta$, we let $I(\theta, T) := \{i \in I \,|\, \theta_i \in T_i\}$ denote the set of agents who can receive the good at that state when their types are in T.[13] We then let

$$\phi(\theta, T) := \min_{H \in \mathcal{P}(I(\theta,T))} \sum_{G \in H} C_G$$

be the maximal number of units the agents in $I(\theta, T)$ can receive in state θ, and let

$$H(\theta, T) \in \arg \min_{H \in \mathcal{P}(I(\theta,T))} \sum_{G \in H} C_G$$

be an associated cover, which we call a **minimal cover**.[14]

In terms of our network, $\phi(\theta, T)$ represents the maximal flow the demand nodes $T^D = T_1 \cup \ldots \cup T_{|I|}$ (those corresponding to the profile of type sets T) can receive in state θ—more precisely, from the super source via (I, θ) (i.e., the capacity node corresponding to state θ). The minimal cover $H(\theta, T)$ in turn represents the groups of agents whose total capacity limits form a "bottleneck" on this flow.

[13]For instance, in the example of Section 2.2, suppose the state is $(\bar{\theta}_1, \underline{\theta}_2)$. Then, when their types are in $T = (\{\bar{\theta}_1\}, \{\bar{\theta}_2\})$, only agent 1 can receive the good, so $I((\bar{\theta}_1, \underline{\theta}_2), (\{\bar{\theta}_1\}, \{\bar{\theta}_2\})) = \{1\}$.

[14]Going back to the example, the covers of $I((\bar{\theta}_1, \underline{\theta}_2), (\{\bar{\theta}_1\}, \{\bar{\theta}_2\})) = \{1\}$ are $\{1\}$ and $\{1, 2\} = I$. Since $C_1 = 2 < C_I = 3$, the minimal cover is $\{1\}$ and the associated capacity is $C_1 = 2$. Namely, $H((\bar{\theta}_1, \underline{\theta}_2), (\{\bar{\theta}_1\}, \{\bar{\theta}_2\})) = \{1\}$, and $\phi((\bar{\theta}_1, \underline{\theta}_2), (\{\bar{\theta}_1\}, \{\bar{\theta}_2\})) = C_1 = 2$.

Finally, we let $Y(T) := \{\theta \in \Theta | I(\theta, T) \neq \emptyset\} = \bigcup_{i \in I}(T_i \times \Theta_{-i})$ denote the set of states at which at least one agent i with type in T_i can receive the good. This means that the capacity nodes $\{(I, \theta)\}_{\theta \in Y(T)}$ are precisely those that can send flow to demand nodes T^D. We now present our main characterization:

Theorem 2. *Let $q = (q_i : \Theta_i \to [0, C_i])_{i \in I}$ be an interim allocation rule. Then q is the reduced form of an allocation rule that respects the capacity constraints $(\mathcal{H}, \mathcal{C})$, if and only if for all $T = (T_1, \ldots, T_{|I|})$, where $T_i \subset \Theta_i$,*

$$\sum_{i \in I} \sum_{\theta_i \in T_i} q_i(\theta_i) p_i(\theta_i) \leq \sum_{\theta \in Y(T)} \phi(\theta, T) p(\theta). \tag{B}$$

We first illustrate the main idea of the proof. The formal proof can be found in the Appendix.

Illustration of Proof. The key arguments for the proof are: (a) For each $T = (T_1, ..., T_{|I|})$, $T_i \subset \Theta_i$, Gale's condition (2.4) for a certain set $S = S(T)$ is equivalent to (B) for T, and (b) given these conditions, the conditions (2.4) for all other sets S satisfying $S \cap \Theta^D = T^D$ are redundant. We illustrate the construction of the sets $S(T)$ via the example of Section 2.2.

Suppose that $T = (\{\overline{\theta}_1\}, \{\overline{\theta}_2\})$. We then choose $S(T)$ containing demand nodes $T^D = \{\overline{\theta}_1, \overline{\theta}_2\}$ such that the directed edges across from $N \backslash S(T)$ into $S(T)$ are the most binding bottlenecks. To do this, we first observe that only the capacities available in states $Y(T) = \{(\overline{\theta}_1, \overline{\theta}_2), (\overline{\theta}_1, \underline{\theta}_2), (\underline{\theta}_1, \overline{\theta}_2)\}$ can be allocated to agent types in T^D. For each state $\theta \in Y(T)$, we define "roots" given by the nodes associated with the minimal cover of the agents $i \in I(\theta, T)$. At state $(\overline{\theta}_1, \overline{\theta}_2)$, both agents can receive flow at the maximum capacity of $\phi((\overline{\theta}_1, \overline{\theta}_2), T) = C_I = 3$, so the minimum cover is I with associated root $(I, (\overline{\theta}_1, \overline{\theta}_2))$. At state $(\overline{\theta}_1, \underline{\theta}_2)$, only agent 1 can receive flow at the maximum capacity of $\phi((\overline{\theta}_1, \overline{\theta}_2), T) = C_1 = 2$, so the minimum cover is $\{1\}$ with associated root $(\{1\}, (\overline{\theta}_1, \underline{\theta}_2))$. Similarly, at state $(\underline{\theta}_1, \overline{\theta}_2)$, the minimal cover is $\{2\}$ and the associated root is $(\{2\}, (\underline{\theta}_1, \overline{\theta}_2))$ and the capacity is $\phi((\underline{\theta}_1, \overline{\theta}_2), T) = C_2 = 2$. Roots are marked by bold-faced nodes in Figure 3.1. In addition to these three roots, all the nodes that are on the path from these roots to the demand nodes $T^D = \{\overline{\theta}_1, \overline{\theta}_2\}$ are included in $S(T)$. In Figure 3.1, the shaded nodes are the set $S(T)$. Notice that all edges from $N \backslash S(T)$ enter $S(T)$ only via these roots, and these edges (marked as thick arrows in Figure 3.1) form the bottlenecks. The total capacity of the edges that enter the roots for all $\theta \in Y(T)$ is the maximum flow that can reach the set T^D. It equals the sum of $\phi(\theta, T)$'s weighted by $p(\theta)$'s, thus giving the right-side of (2.6). Meanwhile, $d(S(T))$ is simply the total demand at $T^D = \{\overline{\theta}_1, \overline{\theta}_2\}$, which equals the left-side of (2.6). This shows that (2.6) is given by (2.4) for $S = S(T)$.

Furthermore, (2.4) for any $S \neq S(T)$ such that $S \cap \Theta^D = \{\overline{\theta}_1, \overline{\theta}_2\}$ is implied by (2.6). For instance, consider $S = S(T) \backslash \{(I, (\overline{\theta}_1, \overline{\theta}_2))\}$. The demand $d(S)$ is the same as $d(S(T))$, giving rise to the same quantity on the left side of (2.4), but the total capacity of edges that enter S exceeds the total capacity of edges that enter $S(T)$ by $p(\overline{\theta}_1, \overline{\theta}_2) = (C_1 + C_2 - C_I)p(\overline{\theta}_1, \overline{\theta}_2)$, so the right side gets bigger. □

In the standard one-unit auction, $C_i = C_G = 1$ for all $i \in I$ and $G \in \mathcal{H}$, so that $\phi(\theta, T) = 1$ for all $\theta \in Y(\theta)$. Hence, our characterization simplifies to the familiar condition:

Corollary 1 (Border)**.** *In the standard auction, an interim allocation rule q is the reduced form of an allocation rule if and only if for all $T = (T_1, \ldots, T_{|I|})$, where $T_i \subset \Theta_i$,*

$$\sum_{i \in I} \sum_{\theta_i \in T_i} q_i(\theta_i) p_i(\theta) \leq \sum_{\theta \in Y(T)} p(\theta).$$

Remark 1. The restriction that the constraint structure is hierarchical is crucial for our approach. To illustrate the problem that arises with non-hierarchical constraints, let us consider an example with three agents, $I = \{1, 2, 3\}$ and $\mathcal{H} = \{I, G = \{1, 2\}, G' = \{2, 3\}, \{1\}, \{2\}, \{3\}\}$. This constraint structure is not hierarchical because G and G' are not disjoint, and neither set is a subset of the other. In order to impose the capacity constraints for G in a network, the nodes in $A := \{(\{1\}, \theta), (\{2\}, \theta)\}$ would have to be connected to the super-source through a single capacity node (G, θ). Similarly, in order to impose the capacity constraints for G' in the network, the nodes in $B := \{(\{2\}, \theta), (\{3\}, \theta)\}$ would have to be connected to the super-source through a single capacity node (G', θ). Since the constraint structure is not hierarchical, the node sets A and B overlap and it is impossible to impose both constraints simultaneously.

Remark 2. The characterization results in this and the following sections generalize to the case of continuous type distributios. The formal proof uses methods from Border (1991). To be specific, given an interim allocation rule q satisfying a continuous type version of (B), one can construct an approximation by simple functions, i.e., a sequence of finite interim allocation rules that satisfy the continuous type version of (B) and converge to q. Each finite allocation rule satisfies (B) in an appropriately discretized version of the continuous type model and is therefore implementable by Theorem 2. Finally, the set of implementable interim allocation rules is compact (given an appropriate topology). Therefore, the interim allocation rule q, which is the limit of a sequence of implementable allocation rules, is also implementable.

4. Reduction of Constraints

The characterization in the previous section involves $\prod_{i \in I} |2^{\Theta_i}|$ inequalities. Since this number grows very quickly with the cardinalities of the type spaces, the condition is not very tractable. In this section, we derive two reductions that lead to more tractable characterizations. First, we show that if types are independently distributed, it is sufficient to check (B) for the upper contour sets of the interim allocation functions, i.e., set of types whose interim expected allocations are no smaller than a certain threshold. With this reduction, the number of inequalities that we need to check becomes much smaller, $\prod_{i\in I}|\Theta_i|$ at most. Second, we show that when some group(s) of agents are symmetric, and if we restrict attention to group-symmetric reduced forms, it suffices to check (B) only for those $T=(T_1,\dots,T_{|I|})$ for which the T_i's are identical for agents in the same group.

4.1. **Independent Type Distribution.** Consider the situation where agents' types are independently distributed, i.e. $p(\theta)=\prod_{i\in I}p_i(\theta_i)$, $\forall\theta\in\Theta$. The following result shows that it is sufficient to check (B) for upper contour sets T_i.

Theorem 3. *Suppose that the agents' types are independently distributed. Then, q is the reduced form of an allocation rule that respects $(\mathcal{H},\mathcal{C})$, if and only if* (B) *holds for all $T=(T_i)_{i\in I}$, where $T_i=\{\theta_i\in\Theta_i\,|\,q_i(\theta_i)\geq e_i\}$, for some $e_i\in[0,C_i]$.*

Bayesian incentive compatibility requires that interim allocations are monotonic, in which case the theorem entails even simpler conditions. With monotonicity, an upper contour set boils down to an interval of types above a threshold. Hence, we obtain the following familiar characterization in the standard setup:

Corollary 2. *Consider the standard single-unit setup and suppose that each Θ_i is linearly ordered and q_i is nondecreasing. Then, q is the reduced form of an allocation rule if and only if for all $(\theta_1,\cdots,\theta_{|I|})\in\Theta$,*

$$\sum_{i\in N}\sum_{\theta_i'\geq\theta_i}q_i(\theta_i')p_i(\theta_i')\leq 1-\prod_{i\in I}P_i(\theta_i),$$

where $P_i(\cdot)$ is the cdf of $p_i(\cdot)$, i.e. $P_i(\theta_i)=\sum_{\theta_i'<\theta_i}p_i(\theta_i')$.

4.2. **Generalized Symmetric Environments.** In many environments, a set of agents share similar characteristics. For instance, in procurement auctions, the incumbents and entrants form two groups, and those within the same group have more in common in terms of technologies and other factors than those outside that group. In such a circumstance, it makes sense to view the agents within the same group as symmetric. This can be done formally by considering a model in which the agents are partitioned into groups, and those in the same group are ex ante identical. In such a model, it often suffices to search for an optimal mechanism in the class of **group-symmetric** mechanisms, namely those that treat ex-ante identical buyers identically. As will be seen, with such mechanisms, the task of identifying reduced-forms can be reduced even further to checking (B) only for group-symmetric sets T.

To be more specific, suppose that I can be partitioned into subsets, $G_1, \ldots, G_L$. All agents in each non-singleton set (or group) G_ℓ are symmetric in the following sense:[15] First, for all $i, j \in G_\ell$, $\Theta_i = \Theta_j =: \hat{\Theta}_\ell$. Second, p is invariant to the permutation of types for any pair of agents $i, j \in G_\ell$, i.e., $p(\theta_i, \theta_j, \theta_{-ij}) = p(\theta_j, \theta_i, \theta_{-ij})$ for all $\theta_i, \theta_j \in \hat{\Theta}_\ell$ and all $\theta_{-ij} \in \Theta_{-ij}$. This implies that for each group, there exists a marginal distribution $\hat{p}_\ell : \hat{\Theta}_\ell \to [0,1]$, satisfying $p_i(\theta_\ell) = \hat{p}_\ell(\theta_\ell)$ for all $\theta_\ell \in \hat{\Theta}_\ell$ and all $i \in G_\ell$. Note that we do *not* require the type distribution to be independent. Third, agents from the same group have identical capacity constraints; i.e., for all G_ℓ, there is some $\hat{C}_\ell$ with $C_i = \hat{C}_\ell$ for all $i \in G_\ell$. Fourth, we assume that there is no non-singleton set $G \subsetneq G_\ell$ with $G \in \mathcal{H}$, i.e., the sets G_ℓ are groups in the lowest tier of the hierarchy (except for the singleton sets). We do not impose any restrictions on higher tiers of the hierarchy.

We call the environment described so far a *generalized symmetric environment* and establish a reduction of our characterization that applies to group-symmetric reduced forms. Formally, a reduced form is group-symmetric if for each group G_ℓ, there exists an interim allocation rule $\hat{q}_\ell : \hat{\Theta}_\ell \to \mathbb{R}_+$ such that $q_i(\theta_\ell) = \hat{q}_\ell(\theta_\ell)$ for all $i \in G_\ell$ and all $\theta_\ell \in \hat{\Theta}_\ell$.

Theorem 4. *In the generalized symmetric environment, a group-symmetric interim allocation rule q is the reduced form of an allocation rule that respects $(\mathcal{H}, \mathcal{C})$ if and only if* (B) *holds for all T satisfying $T_i = T_j$ for all $i, j \in G_\ell$ and all $\ell = 1, \ldots, L$.*

If types are independently distributed, the reductions in Theorems 3 and 4 can be combined:

Corollary 3. *Suppose the agents' types are independently distributed. Then, in the generalized symmetric environment, q is the reduced form of an allocation rule that respects $(\mathcal{H}, \mathcal{C})$ if and only if* (B) *holds for all group-symmetric T's where each T_i is an upper contour set of q_i.*

The original characterization by Border (1991) and its extension by Mierendorff (2011) *without* capacity constraint are special cases of this corollary.

Remark 3. Suppose that agents from two different groups are ex-ante identical but they face two separate (but identical) capacity constraints. Even in such a case we can repeat the argument in Theorem 4 to show that the set T can be made symmetric for all agents in those groups. To be precise, consider the generalized symmetric environment and suppose that two groups G_1 and G_2, are symmetric in the following sense: (i) $|G_1| = |G_2|$; (ii) $\hat{C}_1 = \hat{C}_2$ and $C_{G_1} = C_{G_2}$; (iii) $\hat{\Theta}_1 = \hat{\Theta}_2$; and (iv) $p(\theta_{G_1}, \theta_{G_2}, \theta_{-(G_1 \cup G_2)}) = p(\theta_{G_2}, \theta_{G_1}, \theta_{-(G_1 \cup G_2)})$. Then, it suffices to check (B) for all $T = (T_1, \ldots, T_{|I|})$ such that $T_i = T_j$ for all $i, j \in G_1 \cup G_2$.

[15]We do not exclude the possibility of singletons but symmetry does not impose any conditions on these sets.

Remark 4. A symmetric interim allocation rule q satisfying the conditions of Theorem 4 may be implemented by an allocation rule Q that is not symmetric. Note, however, that we can uniformly randomize the identities of buyers that belong to the same group G_ℓ before applying the allocation rule Q. We thereby construct a new allocation rule $\hat{Q}$ that is symmetric and has the same reduced form, i.e., it also implements q.

5. Application: Auctions with Budget- and Capacity-Constrained agents

This section applies our characterization of reduced-form auctions to solve a specific problem and in the process illustrates the relevance of the characterization—namely, how formulating mechanisms in reduced form can make an otherwise intractable problem manageable. To begin, suppose that $n \geq 2$ units of a good are to be allocated to a set I of agents. The seller wishes to maximize revenue subject to the constraint that each agent can receive at most m units of the good and cannot pay more than his budget, which is common and equal to w.

This kind of problem may arise in the sale of government-owned assets, such as radio spectrum. A government sells many licenses, but the number of licenses that can be allocated to a single firm is often limited. This may be because firms have diminishing marginal utilities from licenses or limited abilities to profitably operate beyond certain units. Alternatively, a government may limit the number of licenses accruing to a single firm in order to keep the post-assignment market from being too concentrated. The limited budget reflects a firm's limited access to capital markets.[16]

Finally, revenue maximization is a good model of a government with sufficiently high shadow costs of raising public funds. For simplicity, the agents are assumed to be symmetric. The value of each unit of the good is equal to $\overline{\theta}$ with probability $p \in (0,1)$ and $\underline{\theta} \in (0, \overline{\theta})$ with probability $1-p$. We assume $\frac{n}{|I|} < m < n$; namely, the individual quota is binding enough to prevent a sale of all units to one agent but leaves some scope for competition. We also assume that $w > \underline{\theta}\frac{n}{|I|}$, which will ensure that the budget constraint is not binding for a low-type agent. This model is stylized, but contains a new feature—"firm-specific capacity constraints"—which is relevant in practice.

The seller's problem can be formulated in a direct mechanism which specifies (Q, T) : $\{\underline{\theta}, \overline{\theta}\}^{|I|} \to \Delta \times \mathbb{R}^{|I|}$ where $\Delta := \{(x_1, ..., x_{|I|}) \in [0,m]^{|I|} | \sum_{i \in I} x_i \leq n\}$. $Q_i(\theta_i, \theta_{-i})$ is the "fractional" quantity of the good agent i receives and $T_i(\theta_i, \theta_{-i})$ is his expected payment, when the agents report θ. By an argument similar to Remark 4, without loss we can focus on a symmetric mechanism in which $(Q_i(\theta, \theta'), T_i(\theta, \theta')) = (Q_j(\theta, \theta'), T_j(\theta, \theta'))$ for all i, j, $\theta \in \{\underline{\theta}, \overline{\theta}\}$ and $\theta' \in \{\underline{\theta}, \overline{\theta}\}^{|I|-1}$, and which is invariant to permutations of θ'. Let $\mathcal{X}$ be the set of all such mechanisms.

[16]The relevance of budget constraints in spectrum auctions is well-documented. See Salant (1997) or Jehiel and Moldovanu (2003, p. 296).

The problem facing the seller is:

$$\max_{(Q,T)\in\mathcal{X}} \mathbb{E}\left[\sum_{i\in I} T_i(\tilde{\theta}_i, \tilde{\theta}_{-i})\right]$$

subject to

$$\theta_i\mathbb{E}_{\tilde{\theta}_{-i}}[Q_i(\theta_i,\tilde{\theta}_{-i})]-\mathbb{E}_{\tilde{\theta}_{-i}}[T_i(\theta_i,\tilde{\theta}_{-i})] \geq \theta_i\mathbb{E}_{\tilde{\theta}_{-i}}[Q_i(\theta_i',\tilde{\theta}_{-i})]-\mathbb{E}_{\tilde{\theta}_{-i}}[T_i(\theta_i',\tilde{\theta}_{-i})],\ \forall\theta_i,\theta_i' \in \{\underline{\theta},\overline{\theta}\}, i\in I, \tag{IC}$$

$$\theta_i\mathbb{E}_{\tilde{\theta}_{-i}}[Q_i(\theta_i,\tilde{\theta}_{-i})] - \mathbb{E}_{\tilde{\theta}_{-i}}[T_i(\theta_i,\tilde{\theta}_{-i})] \geq 0, \forall\theta_i \in \{\underline{\theta},\overline{\theta}\}, \forall i \in I, \tag{IR}$$

$$T_i(\theta_i,\theta_{-i}) \leq w, \forall\theta_i \in \{\underline{\theta},\overline{\theta}\}, \forall\theta_{-i} \in \{\underline{\theta},\overline{\theta}\}^{|I|-1}, \forall i \in I. \tag{BC}$$

The first two constraints are standard, requiring truthful reporting as a Bayes-Nash equilibrium and voluntary participation of the agents. The last constraint reflects the agent's budget constraint. In the current context, the presence of this constraint makes the classic Myerson approach inapplicable, since it is difficult to identify an optimal ex-post allocation point-wise, that satisfies this constraint. Instead, the problem is made tractable via a reduced-form auction approach, i.e., by searching for the optimal mechanism directly in "interim" allocations.

To begin, define interim quantity $q_i(\theta_i)$ and interim payment $t_i(\theta_i)$, respectively. Given symmetry, let $\underline{q} := q_i(\underline{\theta}), \overline{q} := q_i(\overline{\theta}), \underline{t} := t_i(\underline{\theta}), \overline{t} := t_i(\overline{\theta}), \forall i \in I$. We can then express the objective as well as (IC) and (IR) using these variables. In particular, as is well known, the incentive constraint implies monotonicity:

$$\overline{q} \geq \underline{q}. \tag{M}$$

and given monotonicity, the incentive constraint binds only for the high type and individual rationality binds only for the low type:

$$\overline{\theta}\overline{q} - \overline{t} = \overline{\theta}\underline{q} - \underline{t}, \tag{IC'}$$

$$\underline{\theta}\underline{q} - \underline{t} = 0. \tag{IR'}$$

Hence, (IC) and (IR) can be replaced by $(M), (IC')$ and (IR'). Further, we can replace (BC) by

$$\overline{t} \leq w. \tag{BC'}$$

This can be seen as follows. First, $\overline{t} \leq w$ is necessary for (BC) since, if $T_i(\cdot,\cdot) \leq w$, we must have $\overline{t} = \mathbb{E}_{\tilde{\theta}}[T_i(\overline{\theta},\tilde{\theta})] \leq w$. Second, (BC') is sufficient since we can simply set $T_i(\underline{\theta},\cdot) \equiv \underline{t}$ and $T_i(\overline{\theta},\cdot) \equiv \overline{t}$, which preserve all other constraints and (BC) given that (M) holds.

Finally, the interim allocations must be "reduced-form auctions," or implementable. Corollary 3 implies that $(\underline{q},\overline{q})$, $\overline{q} \geq \underline{q}$, is a reduced-form auction if and only if

$$|I| \cdot (\overline{q}p + \underline{q}(1-p)) \leq n \tag{B1}$$

$$|I| \cdot \overline{q}p \leq \sum_{k=1}^{|I|} (\min\{n, km\}) \binom{|I|}{k} p^k (1-p)^{|I|-k}. \tag{B2}$$

The first condition implies that the quantity assigned to the agents should not exceed the total supply. (The individual cap of m is not binding since $m > \frac{n}{|I|}$). The second condition requires that the expected quantity assigned to high-type agents does not exceed the expected capacity limits they face.[17] (Notice the expression $\min\{\cdot,\cdot\}$ above corresponds to $\phi(\theta, T)$ in our general characterization.) The feasible set is depicted as the dark shaded area in Figure 5.1. For later use, let $\overline{q}^M$ be the highest $\overline{q}$ that satisfies $(B2)$, and let $\underline{q}^M$ be the highest $\underline{q}$ satisfying $(B1)$ given $\overline{q} = \overline{q}^M$. Let $\eta(\overline{q})$ be the highest $\underline{q}$ satisfying $(B1)$ for any $\overline{q} \in [\frac{n}{|I|}, \overline{q}^M]$.

Summarizing the results so far, the problem $[P]$ simplifies to:

$$[P'] \qquad \max_{(\underline{q},\overline{q})} |I| \cdot (p\overline{t} + (1-p)\underline{t})$$

subject to

$$(M), (IC'), (IR'), (BC'), (B1), (B2).$$

Substituting (IC') and (IR'), $[P']$ in turn simplifies to

$$[P''] \qquad \max_{(\underline{q},\overline{q})} |I| \cdot (p\overline{\theta}\overline{q} + (1-p)J(\underline{\theta})\underline{q})$$

subject to

$$(M), (B1), (B2) \text{ and}$$

$$\overline{\theta}\overline{q} - (\overline{\theta} - \underline{\theta})\underline{q} \leq w, \tag{BC''}$$

where $J(\underline{\theta}) := \underline{\theta} - \frac{p}{1-p}(\overline{\theta} - \underline{\theta})$.

Problem $[P'']$ is a simple linear program. In particular, the budget constraint is now expressed in terms of interim allocations. Its optimal solution can be characterized in a straightforward manner. The optimal interim allocation rule $(\underline{q}^*, \overline{q}^*)$ is described as follows.

(1) If $w \geq \overline{\theta}\overline{q}^M$ and $p > \underline{\theta}/\overline{\theta}$, then $(\underline{q}^*, \overline{q}^*) = (0, \overline{q}^M)$, (a point denoted by A in Figure 5.1). In this case, the budget w is sufficiently large so that it is never binding; in Figure 5.1, the budget constraint line representing (BC'') (not drawn) would lie above the point A. The condition $p > \underline{\theta}/\overline{\theta}$ means that the seller finds it optimal to exclude the low type altogether. ($p > \underline{\theta}/\overline{\theta} \iff J(\underline{\theta}) < 0$.)

(2) If $w \geq \overline{\theta}\overline{q}^M - (\overline{\theta} - \underline{\theta})\underline{q}^M$ and $p \leq \underline{\theta}/\overline{\theta}$, then $(\underline{q}^*, \overline{q}^*) = (\underline{q}^M, \overline{q}^M)$, (a point denoted by C in Figure 5.1).[18] In this case, the budget w is large enough so that the line

[17] There are $k \geq 1$ high type agents with probability $\binom{|I|}{k} p^k (1-p)^{|I|-k}$ and each high type agent can never get more than m units.

[18] We have $\overline{q}^M \geq \underline{q}^M$ since

$$\overline{q}^M = \frac{1}{p} \sum_{k=1}^{|I|} \left(\min\left\{\frac{n}{|I|}, \frac{km}{|I|}\right\}\right) \binom{|I|}{k} p^k (1-p)^{|I|-k}$$

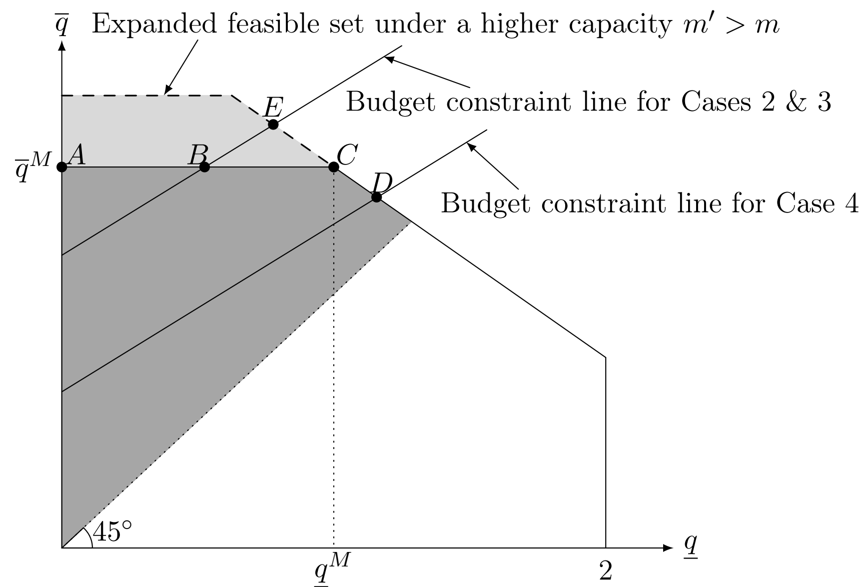

FIGURE 5.1

corresponding to (BC'') would lie above the point C. The condition $p \leq \underline{\theta}/\overline{\theta}$ means that the seller wishes to maximize both $\overline{q}$ and $\underline{q}$.

(3) If $\overline{\theta}\overline{q}^M > w \geq \overline{\theta}\overline{q}^M - (\overline{\theta} - \underline{\theta})\underline{q}^M$ and $p > \underline{\theta}/\overline{\theta}$, then $(\underline{q}^*, \overline{q}^*) = \left(\frac{\overline{\theta}\overline{q}^M - w}{\overline{\theta} - \underline{\theta}}, \overline{q}^M\right)$ (a point denoted by B in Figure 5.1).[19] The condition $p > \underline{\theta}/\overline{\theta}$ means that the seller would exclude the low type, absent the budget constraint. This latter constraint is binding, however, which limits the payment she can extract from the high type via exclusion. This lessens the degree of exclusion of the low type.

(4) If $\overline{\theta}\overline{q}^M - (\overline{\theta} - \underline{\theta})\underline{q}^M > w \geq \underline{\theta}\frac{n}{|I|}$, then $(\underline{q}^*, \overline{q}^*)$ solves the equations $\underline{q} = \eta(\overline{q})$ and $\overline{\theta}\overline{q} - (\overline{\theta} - \underline{\theta})\underline{q} = w$.[20] This point is denoted by D in Figure 5.1. In this case, $\overline{q}^* < \overline{q}^M$, so the seller sells less to the high type and more to the low type than under the efficient allocation.

$$= \sum_{j=0}^{|I|-1} \left(\min\left\{\frac{n}{j+1}, m\right\}\right) \binom{|I|-1}{j} p^j (1-p)^{|I|-1-j}$$
$$\geq \sum_{j=0}^{|I|-1} \left(\min\left\{\frac{n}{|I|}, m\right\}\right) \binom{|I|-1}{j} p^j (1-p)^{|I|-1-j}$$
$$= \frac{n}{|I|} = p\overline{q}^M + (1-p)\underline{q}^M.$$

[19]It is easily verified that the budget line is always steeper than the iso-profit line. Therefore $\overline{q}^* = \overline{q}^M$ is optimal (rather than $\overline{q}^* = w/\overline{\theta}$).

[20]$\overline{q}^* < \underline{q}^*$ implies $w < \underline{\theta}\overline{q}^M = \underline{\theta}\overline{q}^* < \underline{\theta}\underline{q}^*$, which violates the budget constraint. Hence $\overline{q}^* \geq \underline{q}^*$.

Two observations are worth making. First, the degree of exclusion is lessened by the presence of budget constraints.[21] The reason is that the rent that the seller can extract from the high type by excluding the low type is limited by the budget constraint green. Second, the optimal mechanism involves "random" allocation in Case 4. That is, the low type gets some units with positive probability even when the high type is not fully assigned up to its capacity; so the optimal mechanism "distorts" the high-type's allocation downward, which stands in contrast to the standard problem absent budget constraints. Indeed, these are some features of optimal mechanisms when agents are financially constrained (see Laffont and Robert (1996), Che and Gale (2000), Maskin (2000), and Pai and Vohra (2008)). The current analysis, enabled by our new methodology, shows that they reemerge in the presence of the "firm-specific capacity" constraint.

More interestingly, our new characterization allows us to study the effect of the firm's capacity constraint. Suppose the firm capacity rises from m to $m' > m$. This means that the feasible set expands to the light shaded area as depicted in Figure 5.1. Evidently, no changes occur when the initial solution was at D in Case 4. Changes occur in the other three cases. In all these cases, not surprisingly, $\overline{q}^*$ rises since more units can be sold to the high type now. It is also not surprising that in Case 2 this results in fewer units being sold to the low type; i.e., $\underline{q}^*$ falls as we move from C to E. Interestingly, more units can be sold to the low type as well; in Case 3 (and possibly in Case 1), both $\overline{q}^*$ and $\underline{q}^*$ rise (for Case 3, this is depicted as a shift from B to E in the figure). The reason has to do with the binding budget constraint in this case; selling more to the high type reduces the benefit from "partially" excluding the low type further, making it profitable to sell more to the low type. In this case, efficiency improves more than can be accounted for by the increased capacity.

Appendix A. Omitted Proofs

Proof of Lemma 1. ("If" part). Fix an interim allocation rule q, and let d be the associated demand. Suppose there exists a feasible flow f for demand d in the network. Then,

$$\sum_{\theta_{-i}\in\Theta_{-i}} f((i,(\theta_i,\theta_{-i})),\theta_i) \geq d(\theta_i) = q_i(\theta_i)p_i(\theta_i).$$

Then, for each θ_i and θ_{-i}, there exists $\tilde{f}((i,(\theta_i,\theta_{-i})),\theta_i) \in [0, f((i,(\theta_i,\theta_{-i})),\theta_i)]$ such that

$$\sum_{\theta_{-i}\in\Theta_{-i}} \tilde{f}((i,(\theta_i,\theta_{-i})),\theta_i) = d(\theta_i) = q_i(\theta_i)p_i(\theta_i).$$

[21] Suppose $J(\underline{\theta}) < 0$. When w is sufficiently high, the optimal solution is in Case 1. But as w falls, the optimal solution shifts to Case 2 initially and finally to Case 4 when w falls even further. In the process, $\underline{q}^*$ increases.

We can extend $\tilde{f}$ to the entire set E of edges: For each capacity node (G, θ), define the flow on edge $(n, (G, \theta)) \in E$ via the flow conservation law: $\tilde{f}(n, (G, \theta)) := \sum_{i \in G} \tilde{f}((i, \theta), \theta_i)$. Clearly, $\tilde{f}$ constructed in this way is a feasible flow for demand d.

Now define the allocation rule Q given by:

$$Q_i(\theta) := \frac{\tilde{f}((i, \theta), \theta_i)}{p(\theta)}.$$

This rule respects $(\mathcal{H}, \mathcal{C})$ since, for any $G \in \mathcal{H}$,

$$\sum_{i \in G} Q_i(\theta) = \sum_{i \in G} \frac{\tilde{f}((i, \theta), \theta_i)}{p(\theta)} = \frac{\tilde{f}((G', \theta), (G, \theta))}{p(\theta)} \leq \frac{c((G', \theta), (G, \theta))}{p(\theta)} = C_G,$$

where G' is the predecessor of G. Further, Q implements q since, for each θ_i,

$$q_i(\theta_i) = \frac{\sum_{\theta_{-i} \in \Theta_{-i}} \tilde{f}((i, (\theta_i, \theta_{-i})), \theta_i)}{p_i(\theta_i)} = \sum_{\theta_{-i} \in \Theta_{-i}} Q_i(\theta_i, \theta_{-i}) \frac{p(\theta_i, \theta_{-i})}{p_i(\theta_i)} = \sum_{\theta_{-i} \in \Theta_{-i}} Q_i(\theta_i, \theta_{-i}) p_{-i}(\theta_{-i} | \theta_i),$$

satisfying (3.1).

("Only if" part) Suppose an interim allocation rule q is implementable. Then, there exists an allocation rule Q that respects $(\mathcal{H}, \mathcal{C})$ and satisfies (3.1). Define a mapping $f : E \to \mathbb{R}_+^{|E|}$ given by

$$f((i, (\theta_i, \theta_{-i})), \theta_i) := Q_i(\theta_i, \theta_{-i}) p(\theta_i, \theta_{-i}), \forall i \in I, \forall \theta_i, \forall \theta_{-i},$$

and

$$f(n, (G, \theta)) := \sum_{i \in G} Q_i(\theta) p(\theta), \text{ for } (n, (G, \theta)) \in E,$$

which guarantees the flow conservation law. Further, since Q respects $(\mathcal{H}, \mathcal{C})$, for any n with $(n, (G, \theta)) \in E$,

$$f(n, (G, \theta)) = \sum_{i \in G} Q_i(\theta) p(\theta) \leq C_G p(\theta) = c(n, (G, \theta)).$$

Hence, f is a flow on (N, E, c). Finally, since Q satisfies (3.1), for each θ_i,

$$\sum_{\theta_{-i} \in \Theta_{-i}} \tilde{f}((\theta_i, \theta_{-i}), \theta_i) = q_i(\theta_i) p_i(\theta_i) = d(\theta_i),$$

so f is a feasible flow for d. □

Proof of Theorem 2. *Necessity.*—For given T, we construct a node set $S(T)$ such that Gale's condition (2.4) for $S = S(T)$ implies (B). The set of demand nodes contained in $S(T)$ is given by $T^D := \bigcup_{i \in I} T_i$. Next, for each $\theta \in Y(T)$, $S(T)$ contains capacity nodes called **roots** for (θ, T) defined by $R(\theta, T) := \{(G, \theta) | G \in H(\theta, T)\}$. In words, roots for (θ, T) consist of capacity nodes (G, θ) associated with the minimal cover $H(\theta, T)$.[22] Finally, it contains

[22]We can select any minimal cover if the minimal cover is not unique.

every capacity node (G,θ) that lies on a directed path from some node $(G',\theta) \in R(\theta,T)$ to some demand node $\theta_i \in T^D$:[23]

$$(G,\theta) \in S(T) \iff \left\{\exists G' \in R(\theta,T) : G \subset G' \text{ and } \theta_i \in T^D \text{ for some } i \in G\right\}.$$

Since $S(T)$ contains exactly the demand nodes in T^D and does not contain the super-source, its demand is given by

$$d(S(T)) = \sum_{\theta_i \in T^D} d(\theta_i) = \sum_{i \in I} \sum_{\theta_i \in T_i} q_i(\theta_i) p_i(\theta), \tag{A.1}$$

which equals the left side of (B).

Meanwhile,

$$\begin{aligned} c(N \backslash S(T), S(T)) &= c\left(N \backslash S(T), \bigcup_{\theta \in Y(T)} R(\theta,T)\right) \\ &= \sum_{\theta \in Y(T)} \left[\sum_{(G,\theta) \in R(\theta,T)} C_G\right] p(\theta) \\ &= \sum_{\theta \in Y(T)} \left[\sum_{G \in H(\theta,T)} C_G\right] p(\theta) \\ &= \sum_{\theta \in Y(T)} \phi(\theta,T) p(\theta). \end{aligned} \tag{A.2}$$

The first equality holds since all edges from $N \backslash S(T)$ enter $S(T)$ only through $\bigcup_{\theta \in Y(T)} R(\theta,T)$; the second is obtained by summing up the capacities of those edges that enter from outside $S(T)$ into the roots; the third follows from the definition of roots, and the last equality follows from the fact that $\phi(\theta,T)$ is the capacity associated with the minimal cover $H(\theta,T)$. Note that the last line of (A.2) equals the right side of (B).

Therefore, given (A.1) and (A.2), (B) follows from (2.4) for $S = S(T)$.

Sufficiency.—We now prove that (2.4) is redundant unless $S = S(T)$ for some $T = (T_1, \ldots, T_{|I|})$. Fix an arbitrary $S \subset N$. If $\sigma \in S$, then $d(S) = -\infty$. Likewise, if $S \cap \Theta^D = \emptyset$, then $d(S) \leq 0$. In these two cases, (2.4) is satisfied for all interim allocation rules. Hence, it suffices to consider the case where $S \cap \Theta^D = T^D = \bigcup_{i \in I} T_i \neq \emptyset$ for some $T = (T_1, ..., T_I)$ and $\sigma \notin S$. Furthermore, we can restrict attention to sets S for which $(i, (\theta_i, \theta_{-i})) \in S$ for all $\theta_i \in S$. Otherwise $S(N \backslash S, S) = \infty$ since $c((i,\theta), \theta_i) = \infty$ by our construction, so (2.4) always holds.

[23]Excluding some (G,θ) on the path from (G',θ) to θ_i would introduce an additional edge from $N \backslash S$ to S. Therefore, if we do not include (G,θ) in S, the capacity $c(N \setminus S, S)$ would not be the least upper bound for the flow from (I,θ) to the nodes $\{\theta_i \,|\, i \in I(\theta,T)\}$.

For such a set S, define "roots" $R_S(\theta) \subset S$ for each $\theta \in Y(T)$ such that: $(G,\theta) \in R_S(\theta)$ if and only if either $G = I$ or $G \neq I$ and $(G',\theta) \notin S$, where G' is the predecessor of G. Remember that for every θ and every $i \in I(\theta,T)$, we have $(i,\theta) \in S$. Since $\sigma \notin S$, there must be some (G,θ) on the path from σ to (i,θ) for each $i \in I(\theta,T)$ such that $i \in G$ and $(G,\theta) \in R_S(\theta)$ (where we allow $(G,\theta) = (i,\theta)$). This means that the set $H_S(\theta) := \{G \in \mathcal{H} \,|\, (G,\theta) \in R_S(\theta)\}$ covers $I(\theta,T)$. But if $H_S(\theta) \in \mathcal{P}(I(\theta,T))$, we have

$$\sum_{(G,\theta)\in R_S(\theta)} C_G = \sum_{G\in H_S(\theta)} C_G \geq \min_{H\in\mathcal{P}(I(\theta,T))} \sum_{G\in H} C_G = \phi(\theta,T). \tag{A.3}$$

Multiplying (A.3) by $p(\theta)$ and summing up over all $\theta \in Y(\theta)$, we obtain $c(N\backslash S,S)$. Hence, we have

$$\begin{aligned} c(N\backslash S,S) - d(S) &= c(N\backslash S,S) - d(S\cap\Theta^D) \\ &\geq \sum_{\theta\in Y(T)} \phi(\theta,T)p(\theta) - d(S\cap\Theta^D) \\ &= \sum_{\theta\in Y(T)} \phi(\theta,T)p(\theta) - \sum_{i\in I}\sum_{\theta_i\in T_i} q_i(\theta_i)p_i(\theta) \geq 0. \end{aligned}$$

The first equality holds since S does not contain the super-source. The first inequality holds by (A.3). The second equality holds since $d(S\cap\Theta^D) = d(T^D) = \sum_{i\in I}\sum_{\theta_i\in T_i} q_i(\theta_i)p_i(\theta)$. The last inequality follows from (B). The above string of inequalities gives the sufficiency. □

Proof of Theorem 3: The necessity part is obvious. To establish sufficiency, we fix one agent i and arbitrary type sets T_{-i}. Then, for any $T_i \subset \Theta_i$, (B) becomes

$$\begin{aligned} \sum_{j\in I}\sum_{\theta_j\in T_j} q_j(\theta_j)p_j(\theta) &\leq \sum_{\theta\in Y(T)} \phi(\theta,T)p(\theta), \\ &= \underbrace{\sum_{\theta\in\bigcup_{j\neq i}(T_j\times\Theta_{-j})} \phi(\theta,T)p(\theta)}_{=\alpha_i(T)} + \sum_{\theta\in(T_i\times\Theta_{-i})\backslash\bigcup_{j\neq i}(T_j\times\Theta_{-j})} \phi(\theta,T)p(\theta) \\ &= \alpha_i(T) + \sum_{\theta_i\in T_i} \beta_i(T_{-i})\, C_i p_i(\theta_i), \end{aligned}$$

where $\beta_i(T_{-i}) = \prod_{j\neq i}(1 - p_j(T_j))$ and C_i is the capacity constraint for agent i. To see that the second equality in the above equation holds, note that in the second sum in the second line, we have $\theta_i \in T_i$ and $\theta_j \notin T_j$ for all $j \neq i$. But this implies that $\phi(\theta,T) = C_i$, independent of θ_{-i}. Using independence of the type distribution, we get

$$\sum_{\theta\in(T_i\times\Theta_{-i})\backslash\bigcup_{j\neq i}(T_j\times\Theta_{-j})} C_i p(\theta) = \sum_{\theta_i\in T_i} \beta_i(T_{-i})\, C_i\, p_i(\theta_i).$$

We now rewrite (B) as

$$\Phi(T_i, T_{-i}) := \sum_{\theta_i \in T_i} (q_i(\theta_i) - \beta_i(T_{-i})\, C_i)\, p_i(\theta_i) - \alpha_i(T) \leq -\sum_{j \neq i} \sum_{\theta_j \in T_j} q_j(\theta_j) p_j(\theta_j). \quad \text{(A.4)}$$

For the proof, it will then suffice to show that for given T_{-i}, $\Phi(T_i, T_{-i})$ is maximized by a set T_i that is an upper contour set of q_i.[24]

To begin, we establish the following property of $\alpha_i(\cdot)$.

Claim 1. *For any set T_i and any $\tilde{\theta}_i \in T_i$, let $\tilde{T}_i = T_i \backslash \{\tilde{\theta}_i\}$ and $\tilde{T} = (\tilde{T}_i, T_{-i})$. Then, there is some $\gamma_i(T_{-i}) \geq 0$ such that*

$$\alpha_i(T) - \alpha_i(\tilde{T}) = \gamma_i(T_{-i})\, p_i(\tilde{\theta}_i).$$

Proof. Using the definition of $\alpha_i(\cdot)$, we have

$$\alpha_i(T) - \alpha_i(\tilde{T}) = \sum_{\theta \in \bigcup_{j \neq i}(T_j \times \Theta_{-j})} \big[\phi(\theta, T) - \phi(\theta, \tilde{T})\big] p(\theta).$$

If θ is such that $\theta_i \neq \tilde{\theta}_i$, $\mathcal{P}(\theta, T) = \mathcal{P}(\theta, \tilde{T})$ because $I(\theta, T) = I(\theta, \tilde{T})$. This implies that $\phi(\theta, T) - \phi(\theta, \tilde{T}) = 0$ if $\theta_i \neq \tilde{\theta}_i$ and hence

$$\begin{aligned} \alpha_i(T) - \alpha_i(\tilde{T}) &= \sum_{\theta \in \bigcup_{j \neq i}(T_j \times \{\tilde{\theta}_i\} \times \Theta_{-ij})} \big[\phi(\theta, T) - \phi(\theta, \tilde{T})\big] p(\theta) \\ &= \underbrace{\left(\sum_{\theta_{-i} \in \bigcup_{j \neq i}(T_j \times \Theta_{-ij})} [\phi((\tilde{\theta}_i, \theta_{-i}), T) - \phi((\tilde{\theta}_i, \theta_{-i}), \tilde{T})] p_{-i}(\theta_{-i}) \right)}_{=:\gamma_i(T_{-i})} p_i(\tilde{\theta}_i), \quad \text{(A.5)} \end{aligned}$$

We now argue that the expression in the large parentheses is independent of T_i. It suffices to show that $\phi((\tilde{\theta}_i, \theta_{-i}), T) - \phi((\tilde{\theta}_i, \theta_{-i}), \tilde{T})$ is independent of T_i and $\tilde{\theta}_i$. Observe that $\mathcal{P}(\theta, T)$ depends on θ and T only though $I(\theta, T)$. But $I((\tilde{\theta}_i, \theta_{-i}), T) = \{j \neq i \,|\, \theta_j \in T_j\} \cup \{i\}$ because $\tilde{\theta}_i \in T_i$, independent of the particular choice of $\tilde{\theta}_i$ and T_i. Similarly, and $I((\tilde{\theta}_i, \theta_{-i}), \tilde{T}) = I((\tilde{\theta}_i, \theta_{-i}), T) \backslash \{i\}$ because $\tilde{\theta}_i \notin \tilde{T}_i$, independent of the particular choice of $\tilde{\theta}_i$ and T_i. This implies that $\mathcal{P}((\tilde{\theta}_i, \theta_{-i}), T)$ and $\mathcal{P}((\tilde{\theta}_i, \theta_{-i}), \tilde{T})$ and thus $\phi((\tilde{\theta}_i, \theta_{-i}), T)$ and $\phi(\tilde{\theta}_i, \theta_{-i}), \tilde{T})$ are independent of T_i and $\tilde{\theta}_i$. □

The claim implies that

$$\Phi(T_i, T_{-i}) = \sum_{\theta_i \in T_i} [q_i(\theta_i) - \beta_i(T_{-i}) C_i - \gamma_i(T_{-i})]\, p(\theta_i) - \alpha_i(\emptyset, T_{-i}).$$

Obviously, this expression is maximized by the upper contour set $T_i = \{\theta_i \in \Theta_i | q_i(\theta_i) \geq \beta_i(T_{-i}) C_i + \gamma_i(T_{-i})\}$. □

[24]The original idea of this proof is from Theorem 4 in Gutmann et al. (1991).

Proof of Theorem 4: Recalling that for any given set $I' \subset I$, $\mathcal{P}(I')$ is the set of all covers of I', let us begin by introducing some notation:[25]

$$\tilde{\phi}(I') := \min_{H \in \mathcal{P}(I')} \sum_{G \in H} C_G \quad \text{and} \quad \tilde{H}(I') := \arg \min_{H \in \mathcal{P}(I')} \sum_{G \in H} C_G.$$

We first establish a couple of properties of the function $\tilde{\phi}$:

Claim 2. *For any $I' \subset I$, and any $\ell \in 1, \ldots, L$,*

$$\tilde{\phi}(I' \cup \{i\}) = \tilde{\phi}(I' \cup \{j\}), \forall i, j \in G_\ell \backslash I' \tag{A.6}$$

$$\tilde{\phi}(I' \cup \{i\}) - \tilde{\phi}(I') \geq \tilde{\phi}(I' \cup \{i, j\}) - \tilde{\phi}(I' \cup \{i\}), \forall i, j \in G_\ell \backslash I'. \tag{A.7}$$

Proof. Fix any $I' \subset I$ and consider any $i \in G_\ell \backslash I'$. Given the capacity structure, it must be that either (i) $\tilde{H}(I' \cup \{i\}) = \tilde{H}(I') \cup \{\{i\}\}$ or (ii) there is some G such that $G_\ell \subset G \in \tilde{H}(I' \cup \{i\})$. Also, in the case (ii), there is some G such that $G_\ell \subset G \in \tilde{H}(I' \cup \{i\})$ only if $G_\ell \subset G \in \tilde{H}(I' \cup \{j\})$ for the same G, and vice versa, which we call property (ii′). We use these properties to prove (A.6) and (A.7).

First, (A.6) is immediate in the case (ii) due to the property (ii′). It is also immediate in the case (i) since then $\tilde{\phi}(I' \cup \{i\}) = \sum_{\tilde{G} \in \tilde{H}(I')} C_{\tilde{G}} + \hat{C}_\ell = \tilde{\phi}(I' \cup \{j\})$.

To prove (A.7), we first note that there are two cases: for any given $i, j \in G_\ell$, either (iii) $\tilde{H}(I' \cup \{i, j\}) = \tilde{H}(I') \cup \{\{i\}, \{j\}\}$ or (iv) there is some G' such that $G_\ell \subset G' \in \tilde{H}(I' \cup \{i, j\})$. Observe that (iii) implies (i).[26] In the case (iii), it is clear that $\tilde{H}(I' \cup \{i\}) = \tilde{H}(I') \cup \{\{i\}\}$ so $\tilde{\phi}(I' \cup \{i\}) - \tilde{\phi}(I') = \hat{C}_\ell = \tilde{\phi}(I' \cup \{i, j\}) - \tilde{\phi}(I' \cup \{i\})$. In the case (iv), we consider two sub cases depending on whether (i) or (ii) holds for $\tilde{H}(I' \cup \{i\})$. In case (ii) holds, it is clear that $G = G'$ (since both sets contain the entire set G_ℓ) and also $\tilde{H}(I' \cup \{i, j\}) \backslash \{G'\} = \tilde{H}(I' \cup \{i\}) \backslash \{G\}$, which implies that the right side of (A.7) is equal to zero so (A.7) holds trivially. In case (i) holds, note that $\tilde{\phi}(I' \cup \{i\}) = \sum_{\tilde{G} \in \tilde{H}(I')} C_{\tilde{G}} + \hat{C}_\ell$. Note also that, due to the definition of $\tilde{\phi}(\cdot)$, the case (iv) can only hold when $\tilde{\phi}(I' \cup \{i, j\}) \leq \sum_{\tilde{G} \in \tilde{H}(I')} C_{\tilde{G}} + 2\hat{C}_\ell$. Thus, $\tilde{\phi}(I' \cup \{i\}) - \tilde{\phi}(I') = \hat{C}_\ell \geq \tilde{\phi}(I' \cup \{i, j\}) - \tilde{\phi}(I' \cup \{i\})$, as desired. □

We establish the desired result by fixing an arbitrary group $G_\ell \subset I$ and assuming (without loss) that $G_\ell = \{1, \cdots, |G_\ell|\}$. Our proof consists of three steps. For the first step, let us define

$$\Psi(T) := \sum_{\theta \in Y(T)} \phi(\theta, T) p(\theta) = \sum_{\theta \in Y(T)} \tilde{\phi}(I(\theta, T)) p(\theta) = \sum_{\theta \in Y(T)} \sum_{G \in \tilde{H}(I(\theta, T))} C_G.$$

---

[25]The functions $\tilde{\phi}$ and $\tilde{H}$ are analogous to ϕ and H, respectively, except that the former are defined on a set I' while the latter on a pair (θ, T): $\phi(\theta, T) = \tilde{\phi}(I(\theta, T))$ and $H(\theta, T) = \tilde{H}(I(\theta, T))$.

[26]If (i) does not hold, there is some G'' with $G_\ell \subset G'' \in \tilde{H}(I' \cup \{i\})$ as in case (ii). Then, it must be that $G_\ell \subset G'' \in \tilde{H}(I' \cup \{i, j\})$, which corresponds to the case (iv).

Note that $\Psi(T)$ corresponds to the ride side of (B) and, due to the group-symmetry, $\Psi(T)$ is invariant to the permutation of sets $(T_i)_{i\in G_\ell}$.

Step 1. *Consider any pair of agents $i, j \in G_\ell$. Then, for any $T_i, T_j \subset \hat{\Theta}_\ell$ and T_{-ij},*

$$2\Psi(T_i, T_j, T_{-ij}) \geq \Psi(T_i \cup T_j, T_i \cup T_j, T_{-ij}) + \Psi(T_i \cap T_j, T_i \cap T_j, T_{-ij}). \tag{A.8}$$

Proof. To simplify notation, let $I(\theta_{-ij}) := \{h \in I \,|\, \theta_h \in T_h, h \neq i, j\}$. Observe first that for any $h \in I$ and any $X \subset Y \subset \Theta_h$,

$$\Psi(Y, T_{-h}) - \Psi(X, T_{-h}) = \sum_{\theta\in(Y\backslash X)\times\Theta_{-h}} \big[\phi(\theta, (Y, T_{-h})) - \phi(\theta, (X, T_{-h}))\big] p(\theta).$$

Using this, (A.6), and the fact that for any sets X and Y, $(X \cup Y)\backslash X = Y\backslash X$, one can obtain[27]

$$\begin{aligned}
&\Psi(T_i \cup T_j, T_i \cup T_j, T_{-ij}) - \Psi(T_i, T_j, T_{-ij})\\
=&\Psi[(T_i \cup T_j, T_i \cup T_j, T_{-ij}) - \Psi(T_i, T_i \cup T_j, T_{-ij})] + [\Psi(T_i, T_i \cup T_j, T_{-ij}) - \Psi(T_i, T_j, T_{-ij})]\\
=&\sum_{\theta\in(T_j\backslash T_i)\times\Theta_j\times\Theta_{-ij}} \big[\phi(\theta, (T_i \cup T_j, T_i \cup T_j, T_{-ij})) - \phi(\theta, (T_i, T_i \cup T_j, T_{-ij}))\big] p(\theta)\\
&+ \sum_{\theta\in\Theta_i\times(T_i\backslash T_j)\times\Theta_{-ij}} \big[\phi(\theta, (T_i, T_i \cup T_j, T_{-ij})) - \phi(\theta, (T_i, T_j, T_{-ij}))\big] p(\theta)\\
=&\sum_{\theta\in(T_j\backslash T_i)\times(T_i\cup T_j)\times\Theta_{-ij}} \big[\phi(\theta, (T_i \cup T_j, T_i \cup T_j, T_{-ij})) - \phi(\theta, (T_i, T_i \cup T_j, T_{-ij}))\big] p(\theta)\\
&\qquad + \sum_{\theta\in(T_j\backslash T_i)\times(T_i\cup T_j)^c\times\Theta_{-ij}} \big[\phi(\theta, (T_i \cup T_j, T_i \cup T_j, T_{-ij})) - \phi(\theta, (T_i, T_i \cup T_j, T_{-ij}))\big] p(\theta)\\
&+ \sum_{\theta\in T_i\times(T_i\backslash T_j)\times\Theta_{-ij}} \big[\phi(\theta, (T_i, T_i \cup T_j, T_{-ij})) - \phi(\theta, (T_i, T_j, T_{-ij}))\big] p(\theta)\\
&\qquad + \sum_{\theta\in T_i^c\times(T_i\backslash T_j)\times\Theta_{-ij}} \big[\phi(\theta, (T_i, T_i \cup T_j, T_{-ij})) - \phi(\theta, (T_i, T_j, T_{-ij}))\big] p(\theta)\\
=&\sum_{\theta_{-ij}\in\Theta_{-ij}} \big[\tilde{\phi}(I(\theta_{-ij}) \cup \{i, j\}) - \tilde{\phi}(I(\theta_{-ij}) \cup \{i\})\big] \big[p(T_j\backslash T_i, T_i \cup T_j, \theta_{-ij}) + p(T_i, T_i\backslash T_j, \theta_{-ij})\big]\\
&\quad + \sum_{\theta_{-ij}\in\Theta_{-ij}} \big[\tilde{\phi}(I(\theta_{-ij}) \cup \{i\}) - \tilde{\phi}(I(\theta_{-ij}))\big] \big[p(T_j\backslash T_i, (T_i \cup T_j)^c, \theta_{-ij}) + p(T_i^c, T_i\backslash T_j, \theta_{-ij})\big].
\end{aligned}$$

Using a similar derivation, one can also obtain

$$\begin{aligned}
&\Psi(T_i, T_j, T_{-ij}) - \Psi(T_i \cap T_j, T_i \cap T_j, T_{-ij})\\
=&\sum_{\theta_{-ij}\in\Theta_{-ij}} \big[\tilde{\phi}(I(\theta_{-ij}) \cup \{i, j\}) - \tilde{\phi}(I(\theta_{-ij}) \cup \{i\})\big] \big[p(T_i, T_j\backslash T_i, \theta_{-ij}) + p(T_i\backslash T_j, T_i \cap T_j, \theta_{-ij})\big]
\end{aligned}$$

[27] In the derivation below, $p(X, Y, \theta_{-ij})$ for any set $X, Y \subset \hat{\Theta}_\ell$ denotes the probability that agent i and j's types belong to X and Y, respectively, while other agents' type profile is θ_{-ij}.

$$+\sum_{\theta_{-ij}\in\Theta_{-ij}}\left[\tilde{\phi}(I(\theta_{-ij})\cup\{i\})-\tilde{\phi}(I(\theta_{-ij}))\right]\left[p(T_i^c,T_j\backslash T_i,\theta_{-ij})+p(T_i\backslash T_j,(T_i\cap T_j)^c,\theta_{-ij})\right].$$

Combining the derivations so far, we get

$$\begin{aligned}
&\Psi(T_i\cup T_j,T_i\cup T_j,T_{-ij})+\Psi(T_i\cap T_j,T_i\cap T_j,T_{-ij})-2\Psi(T_i,T_j,T_{-ij})\\
=&[\Psi(T_i\cup T_j,T_i\cup T_j,T_{-ij})-\Psi(T_i,T_j,T_{-ij})]-[\Psi(T_i,T_j,T_{-ij})-\Psi(T_i\cap T_j,T_i\cap T_j,T_{-ij})]\\
=&\sum_{\theta_{-ij}\in\Theta_{-ij}}\left[\tilde{\phi}(I(\theta_{-ij})\cup\{i,j\})-\tilde{\phi}(I(\theta_{-ij})\cup\{i\})\right]\underbrace{\begin{bmatrix}p(T_j\backslash T_i,T_j\cup T_i,\theta_{-ij})+p(T_i,T_i\backslash T_j,\theta_{-ij})\\-p(T_i,T_j\backslash T_i,\theta_{-ij})-p(T_i\backslash T_j,T_i\cap T_j,\theta_{-ij})\end{bmatrix}}_{=:A}\\
&+\sum_{\theta_{-ij}\in\Theta_{-ij}}\left[\tilde{\phi}(I(\theta_{-ij})\cup\{i\})-\tilde{\phi}(I(\theta_{-ij}))\right]\underbrace{\begin{bmatrix}p(T_j\backslash T_i,(T_i\cup T_j)^c,\theta_{-ij})+p(T_i^c,T_i\backslash T_j,\theta_{-ij})\\-p(T_i^c,T_j\backslash T_i,\theta_{-ij})-p(T_i\backslash T_j,(T_i\cap T_j)^c,\theta_{-ij})\end{bmatrix}}_{=:B}\\
=&\sum_{\theta_{-ij}\in\Theta_{-ij}}\begin{bmatrix}\tilde{\phi}(I(\theta_{-ij})\cup\{i,j\})+\tilde{\phi}(I(\theta_{-ij}))\\-2\tilde{\phi}(I(\theta_{-ij})\cup\{i\})\end{bmatrix}\left[p(T_i\backslash T_j,T_i\backslash T_j,\theta_{-ij})+p(T_j\backslash T_i,T_j\backslash T_i,\theta_{-ij})\right]\le 0
\end{aligned}$$

as desired, where the third equality follows from the fact that given the invariance of p to the permutation, $A=-B=p(T_i\backslash T_j,T_i\backslash T_j,\theta_{-ij})+p(T_j\backslash T_i,T_j\backslash T_i,\theta_{-ij})$[28] while the inequality follows from (A.7) with $I'=I(\theta_{-ij})$. □

For the next step, let us fix $(T_{k+2},\cdots,T_{|I|})$ and define a function $\widetilde{\Psi}(\cdot)$ as

$$(T_1,\cdots,T_{k+1})\mapsto\widetilde{\Psi}(T_1,\cdots,T_{k+1})=\Psi(T_1,\cdots,T_{k+1},T_{k+2},\cdots,T_{|I|})$$

to simplify notation.

Step 2. *For any sets $T_\ell,T_{k+1}\subset\hat{\Theta}_\ell$,*

$$(k+1)\widetilde{\Psi}(T_\ell^k,T_{k+1})\ge\widetilde{\Psi}((T_\ell\cup T_{k+1})^{k+1})+\widetilde{\Psi}((T_\ell\cap T_{k+1})^{k+1})+(k-1)\widetilde{\Psi}(T_\ell^{k+1}),\qquad\text{(A.9)}$$

where T_ℓ^k, for instance, is an abbreviation of $(\underbrace{T_\ell,\cdots,T_\ell}_{k\ \text{times}})$.

Proof. We prove this by a mathematical induction. Notice that the result in Step 1 takes care of the inequality (A.9) for the case $k=1$. Assuming now that (A.9) holds with k being $k-1$, we show that it also holds for $k\ge 2$. Letting $\overline{T}:=T_\ell\cup T_{k+1}$ and $\underline{T}:=T_\ell\cap T_{k+1}$, note that $\overline{T}\cup T_\ell=\overline{T}$, $\underline{T}\cap T_\ell=\underline{T}$, and $\overline{T}\cap T_\ell=\underline{T}\cup T_\ell=T_\ell$.

[28]To see this, note for instance

$$\begin{aligned}p(T_j\backslash T_i,T_j\cup T_i,\theta_{-ij})-p(T_i,T_j\backslash T_i,\theta_{-ij})&=p(T_j\backslash T_i,T_j\cup T_i,\theta_{-ij})-p(T_j\backslash T_i,T_i,\theta_{-ij})\\&=p(T_j\backslash T_i,(T_j\cup T_i)\backslash T_i,\theta_{-ij})=p(T_j\backslash T_i,T_j\backslash T_i,\theta_{-ij}),\end{aligned}$$

where the first equality follows from the symmetry of $p(\cdot)$.

Observe first

$$2k\widetilde{\Psi}(T_\ell,\overline{T}^k) = 2k\widetilde{\Psi}(T_\ell,\overline{T},\overline{T}^{k-1}) \geq k\widetilde{\Psi}(\overline{T},\overline{T},\overline{T}^{k-1}) + k\widetilde{\Psi}(T_\ell,\underbrace{T_\ell,\overline{T}^{k-1}}_{=\widetilde{T}})$$
$$\geq k\widetilde{\Psi}(\overline{T}^{k+1}) + \widetilde{\Psi}(T_\ell,\overline{T}^k) + \widetilde{\Psi}(T_\ell,T_\ell^k) + (k-2)\widetilde{\Psi}(T_\ell,\overline{T}^k)$$
$$= k\widetilde{\Psi}(\overline{T}^{k+1}) + \widetilde{\Psi}(T_\ell^{k+1}) + (k-1)\widetilde{\Psi}(T_\ell,\overline{T}^k)$$

where the first inequality follows from (A.8) by setting $T_i = T_\ell$ and $T_j = T_{k+1}$ while the second inequality follows from applying the induction hypothesis for $k-1$ to the set profile $\widetilde{T}$. Rearranging this inequality yields

$$(k+1)\widetilde{\Psi}(T_\ell,\overline{T}^k) \geq k\widetilde{\Psi}(\overline{T}^{k+1}) + \widetilde{\Psi}(T_\ell^{k+1}). \tag{A.10}$$

A similar derivation can be employed to yield

$$(k+1)\widetilde{\Psi}(T_\ell,\underline{T}^k) \geq k\widetilde{\Psi}(\underline{T}^{k+1}) + \widetilde{\Psi}(T_\ell^{k+1}). \tag{A.11}$$

Observe now

$$(k+1)k\widetilde{\Psi}(T_\ell^k,T_{k+1}) = (k+1)k\widetilde{\Psi}(T_\ell,\underbrace{T_\ell^{k-1},T_{k+1}}_{=\widehat{T}}) \tag{A.12}$$
$$\geq (k+1)\big[\widetilde{\Psi}(T_\ell,\overline{T}^k) + \widetilde{\Psi}(T_\ell,\underline{T}^k) + (k-2)\widetilde{\Psi}(T_\ell,T_\ell^k)\big]$$
$$\geq k\widetilde{\Psi}(\overline{T}^{k+1}) + k\widetilde{\Psi}(\underline{T}^{k+1}) + 2\widetilde{\Psi}(T_\ell^{k+1}) + (k+1)(k-2)\widetilde{\Psi}(T_\ell^{k+1})$$
$$= k\widetilde{\Psi}(\overline{T}^{k+1}) + k\widetilde{\Psi}(\underline{T}^{k+1}) + k(k-1)\widetilde{\Psi}(T_\ell^{k+1}), \tag{A.13}$$

where the first inequality follows from applying the induction hypothesis for $k-1$ to the set profile $\widehat{T}$ while the second inequality from (A.10) and (A.11). Divide (A.12) and (A.13) by k to get (A.9) □

To state and prove the last step, define $\mathcal{T}_k := \{(T_1,\cdots,T_{|I|})\,|\,T_i = T_j \subset \hat{\Theta}_\ell, \forall i,j \in \{1,\cdots,k\} \subset G_\ell = \{1,2,\cdots,|G_\ell|\}\}$.

Step 3. *For $k,k+1 \in G_\ell$,* (B) *holds for all sets in $\mathcal{T}_k$ if and only if it holds for all sets in $\mathcal{T}_{k+1}$.*

Proof. The only if part is immediate since $\mathcal{T}_{k+1} \subset \mathcal{T}_k$. To show the if part, let $K := \{1,\cdots,k\}$ and $K+1 := \{1,\cdots,k+1\}$ with some abuse of notation. Also define

$$\Lambda(T) := \sum_{i\in I}\sum_{\theta_i\in T_i} q_i(\theta_i)p_i(\theta_i) \;\text{ and }\; \widetilde{\Lambda}(T_{-(K+1)}) := \sum_{i\notin K+1}\sum_{\theta_i\in T_i} q_i(\theta_i)p_i(\theta_i).$$

Notice that $\Lambda(T)$ is the same as the left side of the inequality (B), which can then be simply written as $\Lambda(T) \leq \Psi(T)$. Supposing now that (B) holds for all $T \in \mathcal{T}_{k+1}$, we show that it also holds for any $T \in \mathcal{T}_k$. Observe first that for any $(T_\ell^k, T_{-K}) \in \mathcal{T}_k$,

$$
\begin{aligned}
&(k+1)\Lambda(T_\ell^k, T_{k+1}, T_{-(K+1)}) \\
=&(k+1)\Big[k\sum_{\theta_\ell \in T_\ell} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \sum_{\theta_{k+1}\in T_{k+1}} \hat{q}_\ell(\theta_{k+1})\hat{p}_\ell(\theta_{k+1}) + \widetilde{\Lambda}(T_{-(K+1)})\Big] \\
=&(k+1)\Big[(k-1)\sum_{\theta_\ell \in T_\ell} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \sum_{\theta_\ell\in T_\ell\cup T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \sum_{\theta_\ell\in T_\ell\cap T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \widetilde{\Lambda}(T_{-(K+1)})\Big] \\
=&\Big[(k+1)\sum_{\theta_\ell\in T_\ell\cup T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \widetilde{\Lambda}(T_{-(K+1)})\Big] + \Big[(k+1)\sum_{\theta_\ell\in T_\ell\cap T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \widetilde{\Lambda}(T_{-(K+1)})\Big] \\
&\qquad + (k-1)\Big[(k+1)\sum_{\theta_\ell\in T_\ell} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \widetilde{\Lambda}(T_{-(K+1)})\Big] \\
=&\Lambda((T_\ell\cup T_{k+1})^{k+1}, T_{-(K+1)}) + \Lambda((T_\ell\cap T_{k+1})^{k+1}, T_{-(K+1)}) + (k-1)\Lambda(T_\ell^{k+1}, T_{-(K+1)}),
\end{aligned}
$$

where the second equality holds since

$$
\sum_{\theta_\ell\in T_\ell\cup T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \sum_{\theta_\ell\in T_\ell\cap T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) = \sum_{\theta_\ell\in T_\ell} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell) + \sum_{\theta_\ell\in T_{k+1}} \hat{q}_\ell(\theta_\ell)\hat{p}_\ell(\theta_\ell).
$$

Then, we obtain the desired result since

$$
\begin{aligned}
&(k+1)\Lambda(T_\ell^k, T_{k+1}, T_{-(K+1)}) \\
=&\Lambda((T_\ell\cup T_{k+1})^{k+1}, T_{-(K+1)}) + \Lambda((T_\ell\cap T_{k+1})^{k+1}, T_{-(K+1)}) + (k-1)\Lambda(T_\ell^{k+1}, T_{-(K+1)}) \\
\le&\Psi((T_\ell\cup T_{k+1})^{k+1}, T_{-(K+1)}) + \Psi((T_\ell\cap T_{k+1})^{k+1}, T_{-(K+1)}) + (k-1)\Psi(T_\ell^{k+1}, T_{-(K+1)}) \\
\le&(k+1)\Psi(T_\ell^k, T_{k+1}, T_{-(K+1)}),
\end{aligned}
$$

where the first inequality follows from the assumption that (B) holds for all $T \in \mathcal{T}_{k+1}$, while the second inequality from Step 2 along with the definition of $\widetilde{\Psi}(\cdot)$. □

Notice now that $\mathcal{T}_1$ is the set of all possible T's while $\mathcal{T}_{|G_\ell|}$ is the set of all group-symmetric T's in which T_i's are identical for all agents in G_ℓ. Applying Step 3 recursively, we conclude that (B) holds for all T if and only if it holds for all-group symmetric T. We can apply the same argument to other groups in order to reach the desired conclusion. □

Proof of Corollary 3: Let $\mathcal{T}_U := \{(T_1, \cdots, T_{|I|}) \,|\, \text{each } T_i \text{ is an upper contour subset of } \Theta_i\}$. Then, from Theorem 3, we know that (B) holds for all T if and only if it holds for all $T \in \mathcal{T}_U$. Noting that if X and $Y \subset \hat{\Theta}_\ell$ are upper contour sets, then so are $X \cup Y$ and $X \cap Y$, one can apply the Step 3 in the proof of Theorem 4 to the collections $\mathcal{T}_U \cap \mathcal{T}_k$ and $\mathcal{T}_U \cap \mathcal{T}_{k+1}$ to prove that (B) holds for all $T \in \mathcal{T}_U \cap \mathcal{T}_k$ if and only if it holds for all $T \in \mathcal{T}_U \cap \mathcal{T}_{k+1}$, which gives us the desired result since $\mathcal{T}_U = \mathcal{T}_U \cap \mathcal{T}_1$ and since $\mathcal{T}_U \cap \mathcal{T}_{|G_\ell|}$ is the collection of sets T that are both upper-contour and group-symmetric. □

References

Armstrong, M. (2000): "Optimal Multi-Object Auctions," *Review of Economic Studies*, 67, 455–481.

Asker, J. and E. Cantillon (2010): "Procurement when Price and Quality Matter," *RAND Journal of Economics*, 41, 1–34.

Border, K. C. (1991): "Implementation of Reduced Form Auctions: A Geometric Approach," *Econometrica*, 59, 1175–1187.

——— (2007): "Reduced Form Auctions Revisited," *Economic Theory*, 31, 167–181.

Brusco, S. and G. Lopomo (2002): "Collusion via Signalling in Simultaneous Ascending Bid Auctions with Heterogeneous Objects, with and without Complementarities," *Review of Economics Studies*, 69, 407–436.

Che, Y.-K., D. Condorelli, and J. Kim (2010): "Weak Cartels and Optimal Auctions," Working paper.

Che, Y.-K., W. Dessein, and N. Kartik (2011): "Pandering to Persuade," Working paper, Columbia University.

Che, Y.-K. and I. Gale (1998): "Standard Auctions with Financially Constrained Bidders," *Review of Economic Studies*, 65, 1–21.

——— (2000): "The optimal Mechanism for Selling to a Budget-Constrained Buyer," *Journal of Economic Theory*, 92, 198–233.

Ford, L. R. J. and D. R. Fulkerson (1956): "Maximal Flow Through a Network," *Canadian Journal of Mathematics*, 8, 399–404.

Gale, D. (1957): "A Theorem on Flows in Networks," *Pacific Journal of Mathematics*, 7, 1073–1082.

Gershkov, A., B. Moldovanu, and X. Shi (2011): "Bayesian and Dominant Strategy Implementation Revisited," Working paper, University of Bonn.

Gutmann, S., J. H. B. Kemperman, J. A. Reeds, and L. A. Shepp (1991): "Existence of Probability Measures with Given Marginals," *Annals of Probability*, 19, 1781–1797.

Hart, S. and P. J. Reny (2011): "Implementation of Reduced Form Auctions: A Simple Approach," Hebrew University of Jerusalem, unpublished working paper.

Hörner, J. and L. Samuelson (2011): "Managing Strategic Buyers," *Journal of Political Economy*, 119, 379–425.

Jehiel, P. and B. Moldovanu (2003): "An Economic Perspective on Auctions," *Economic Policy*, 18, 269–308.

Kellerer, H. G. (1961): "Funktionen auf Produkträumen mit vorgegebenen Marginal-Funktionen," *Mathematische Annalen*, 144, 323–344.

Laffont, J.-J. and J. Robert (1996): "Optimal Auctions with Financially Constrained Buyers," *Economics Letters*, 52, 181–186.

Lorentz, G. G. (1949): "A Problem of Plane Measure," *American Journal of Mathematics*, 71, 417–426.

Manelli, A. M. and D. R. Vincent (2010): "Bayesian and Dominant-Strategy Implementation in the Independent Private Values Model," *Econometrica*, 78, 1905–1938.

Maskin, E. and J. Riley (1984): "Optimal Auctions with Risk Averse Buyers," *Econometrica*, 52, 1473–1518.

Maskin, E. S. (2000): "Auctions, Development, and Privatization: Efficient Auctions with Liquidity-Constrained Buyers," *European Economic Review*, 44, 667–681.

Matthews, S. A. (1984): "On the Implementability of Reduced Form Auctions," *Econometrica*, 52, 1519–1522.

McAfee, R. P. and J. McMillan (1992): "Bidding Rings," *American Economic Review*, 82, 579–599.

Mierendorff, K. (2009): "Optimal Dynamic Mechanism Design with Deadlines," University of Bonn, unpulished working paper.

——— (2011): "Asymmetric Reduced Form Auctions," *Economics Letters*, 110, 41–44.

Miralles, A. (2008): "Cardinal Baysian Allocation Mechanisms without Transfers," Working paper, Boston University.

Myerson, R. B. (1981): "Optimal Auction Design," *Mathematics of Operations Reseach*, 6, 58–63.

Pai, M. and R. Vohra (2008): "Optimal Auctions with Financially Constrained Bidders," Northwestern University, unpublished working paper.

Parlane, S. (2001): "Contracting with Capacity Constrained Suppliers," *Economic Theory*, 17, 619–639.

Ryser, H. J. (1957): "Combinatorial Properties of Matrices of Zeros and Ones," *Canadian Journal of Mathematics*, 9, 371–377.

Salant, D. (1997): "Up in the Air: GTE's Experience in the MTA Auction for Personal Communication Services Licenses," *Journal of Economics and Management Strategy*, 6, 549–572.

Strassen, V. (1965): "The Existence of Probability Measures with Given Marginals," *The Annals of Mathematical Statistics*, 36, 423–439.