The Communication Requirements of Efficient Allocations and Supporting Prices*

Noam Nisan† Ilya Segal‡

February 7, 2004

*An earlier version of this paper was circulated under the title "The Communication Complexity of Efficient Allocation Problems." We are grateful to Paul Milgrom, Luis Rayo, Yoav Shoham, Steve Tadelis, Moshe Tenneholtz, and participants of seminars at Harvard-MIT, Princeton, Stanford, Toronto, and the 2002 DIMACS Workshop on Computational Issues in Game Theory and Mechanism Design for useful comments. The first author was supported by a grant from the Israeli Academy of Sciences. The second author was supported by the Guggenheim Foundation and the National Science Foundation, and was hosted by the Institute for Advanced Study in Princeton during his work on the project. Research assistance by Hui Li is gratefully acknowledged.

†Institute of Computer Science, Hebrew University, Jerusalem, Israel, email: noam@cs.huji.ac.il

‡Department of Economics, Stanford University, Stanford, CA 94305; email: ilya.segal@stanford.edu

Abstract

We show that any communication finding a value-maximizing allocation in a private-information economy must also discover supporting prices (in general personalized and nonlinear). In particular, to allocate L indivisible objects, a price must be revealed for each of the $2^L - 1$ bundles. We prove that all monotonic prices must be used, hence exponential communication in L is needed. Furthermore, exponential communication is needed just to ensure a higher share of surplus than that realized by auctioning all objects as a bundle, or even a higher *expected* surplus (for some probability distribution over valuations). When the utilities are submodular, efficiency still requires exponential communication (and fully polynomial approximation is impossible). When the objects are homogeneous, arbitrarily good approximation is obtained with exponentially less communication than exact efficiency.

1 Introduction

We have recently seen great interest in so-called *combinatorial auctions*, designed to allocate L heterogeneous indivisible items among N bidders whose valuations for the different items can be interdependent. Recent important applications include auctions of FCC spectrum licenses and online procurement (see Vohra and de Vries (2002) for an overview). The objective of an auction is to elicit enough information about bidders' preferences so as to realize an efficient or approximately efficient allocation. The mechanism design literature has used the Revelation Principle to ensure the bidders' incentives to reveal their preferences truthfully (e.g., using the Vickrey-Groves-Clarke transfers). However, full revelation of a bidder's preferences requires naming a willingness to pay for each of the $2^L - 1$ bundles of items. Already with $L = 30$, this would involve the communication of more than one billion numbers, which is beyond the capabilities of any human or machine.

Recognition of the communication problem has prompted researchers to propose simpler mechanisms, in which valuations are not fully revealed. For example, in many proposed iterative auction designs, at each stage bidders only need to describe their preferences over a small number of bundles (see, e.g., Parkes (1999), Bikhchandani et al. (2001), Parkes and Ungar (2001), Ausubel and Milgrom (2002), Kwasnica et al. (2002), Zinkevich et al. (2003)). The hope was that such designs could achieve or at least approximate efficiency, while allowing bidders to communicate much less than their complete preferences. This paper demonstrates that the hope is not justified.

We evaluate the communication requirements of allocation problems by establishing the indispensable role of prices in finding efficient allocations in economies with distributed information. Intellectual origins of this idea lie in the early 20th-century debate on central planning alternatives to the market system. Hayek (1945) argued that prices succinctly summarize the knowledge of "particular circumstances of time and place," which is too enormous to be communicated to a central planner. Hurwicz (1960) and Mount and Reiter (1974) formalized Hayek's intuition by showing that in classical con-

vex economies, the Walrasian price mechanism is "informationally efficient," i.e., verifies Pareto efficient allocations with the least amount of communication (as measured by the dimensionality of the message space) among all mechanisms satisfying a continuity restriction. Furthermore, Jordan (1982) showed that the Walrasian mechanism is a unique individually rational mechanism having this property. These results were extended to convex economies with public goods and externalities (e.g., Sato (1981) and Tian (2001,2004)).

However, the existing literature on the "informational efficiency" of price equilibria does not apply to allocation problems in which the agents' preferences may be nonconvex and the goods may be indivisible, such as the combinatorial auction setting. First, the litetature restricts attention to mechanisms that satisfy a continuity restriction, which precludes the communication of discrete allocations. It does not rule out the possibility that mechanisms that do not satisfy this restriction could achieve efficiency without revealing supporting prices. Second, the literature does not define a price equilibrium concept for allocation problems in which a linear-price equilibrium need not exist.[1] Finally, the literature only considers exact efficiency, and does not examine the potential communication savings from allowing approximation.

The present paper addresses these questions in an allocation problem with privately known preferences. We show that in any such problem, *any* communication mechanism that finds surplus-maximizing allocation must also discover a price equilibrium supporting it, which in general involves nonlinear and personalized prices.[2] This holds even if the agents are truthful (follow the prescribed reporting strategies), thus the necessity of discovering prices has nothing to do with the problem of providing agents with appro-

[1] The literature on public-good economies allows personalized ("Lindahl") prices for the public goods, but still restricts them to be linear, and ensures equilibrium existence by assuming convex preferences over divisible public goods.

[2] A related result is obtained by Parkes (2002). He considers the combinatorial auction problem and shows the necessity of revealing supporting prices by those communication languages that verify efficiency while revealing so-called "outcome-independent" information. Our result is more general in that it does not impose any restrictions on the communication language.

priate incentives. The result also implies that the simpler "nondeterministic" problem of *verifying* the efficiency of a proposed allocation is exactly that of announcing supporting prices along with the allocation.

In the classical convex economy one could restrict attention to prices that are Walrasian (i.e., anonymous and linear in consumption), which are much easier to communicate than the agents' preferences. In the combinatorial allocation problem, on the other hand, we show that a huge price space must be used in order to ensure equilibrium existence. Specifically, we show that any possible $\left(2^L - 1\right)$-dimensional price vector listing the prices of all bundles of objects is a unique equilibrium price vector for some valuation profile. Since any efficient mechanism must communicate such a vector, it follows that it must be at least as extensive as a full revelation of one agent's preferences.

Our approach can also be extended to the problem of *approximating* the maximum total surplus within a constant. For this purpose, note that in the "discretized" problem in which the valuations are restricted to be multiples of $\delta > 0$, any misallocation loses at least surplus δ. Therefore, approximating the maximum surplus within less than δ is at least as hard as realizing exact efficiency in the discretized problem. Since the discretized problem can always be solved with finite communication, the relevant measure of the communication burden is the number of transmitted bits. Such discrete problems have been examined in the computer science field of communication complexity, pioneered by Yao (1979) and surveyed in Kushilevitz and Nisan (1997).[3] Since exact efficiency in the discretized problem still requires the communication of (discrete) prices, we are able to show that guaranteeing a better approximation of efficiency than that achieved by auctioning off all objects as a bundle still requires communicating a very large number of bits, which grows exponentially with L.

[3]The general communication complexity problem is to compute a function (in our case, desired allocation) whose inputs (in our case, agents' preferences) are distributed among agents. The communication complexity literature has developed in parallel with, and shares many techniques with, the economic literature on real-valued communication. For more detailed comparisons of the two literatures, see Marschak (1996) and Van Zandt (1999).

The concept of approximation used above required *uniform* approximation of maximum surplus across all states. Similarly, the communication burden was defined as the *maximum* number of bits transmitted across states. Instead of using such "worst-case" measures, one may assume a probability distribution over possible valuations, and ask how many bits must be transmitted *on expectation* to realize a given *expected* surplus. We show that, for *some* joint probability distribution over the agents' valuations, achieving a higher expected surplus than that from the bundled auction still requires communicating an exponential expected number of bits.[4]

These results imply that the only hope to achieve or approximate efficiency without enormous communication is by focusing on cases in which the agents' preferences (or probability distribution over them) are known a priori to lie in a certain class. One example is given by valuations satisfying the "(gross) substitute property" of Kelso and Crawford (1982) and Gul and Stacchetti (1999). With such valuations, a Walrasian equilibrium with L per item prices exists, and as we show here, it can be found with (truly) polynomial communication. However, the substitute property is very restrictive. We show that for the somewhat larger class of "submodular" valuations (i.e., those exhibiting diminishing marginal utility of items), efficiency still requires very extensive communication, and a fast (so-called "fully polynomial") approximation is impossible.

Finally, we consider the case where the items are known to be homogeneous, and so agents only care about the number of objects consumed. This case exhibits a drastic difference between the communication requirements of exact and approximate efficiency. Namely, exact efficiency again requires at least as much communication as a full description of one agent's preferences, which in this case takes L numbers. On the other hand, approximation within any given ε (more generally, fully polynomial approximation) is achieved with only $O(\log L)$ bits. In the setting considered by Calsamiglia (1977), the homogeneous good to be allocated is *divisible*, and exact efficiency requires infinitely-dimensional communication, yet we construct a fully polynomial approximation that

[4]We prove this particular result using a different technique. The proof uses a lower bound on the communication complexity of approximate set packing, which is derived in the Appendix.

allocates the good in small discrete units (provided that the valuations satisfy a weak smoothness condition). Thus, in this particular case an enormous savings in communication can be achieved with only a slight sacrifice in economic efficiency.

The paper is organized as follows. In Section 2 we describe the general allocation problem, the model of communication, and the measures of the communication burden. In Section 3 we characterize efficient communication as that discovering a price equilibrium. In Section 4 we use this characterization to derive a lower bound on the burden of efficient communication. In Section 5 we define the concepts of approximation and relate them to the analysis of discretized problems. In Section 6 we apply the results to the combinatorial allocation problem. In Section 7 we examine the problem with several restricted classes of valuations. Section 8 presents average-case analysis. Section 9 discusses the relation of our results to the computational complexity literature. Section 10 discusses how agents could be given the incentives to obey the suggested communication protocols. Section 11 concludes.

2 The Allocation Problem and Communication

2.1 The allocation problem

Let N be the finite set of agents, and K be the set of allocations. (With a slight abuse of notation, we will use the same letter to denote a set and its cardinality when this causes no confusion.) An agent's valuation assigns real values to all allocations, and is therefore represented with a vector in $\mathbb{R}^K$. The class of possible valuations of agent $i \in N$ is denoted by $U_i \subset \mathbb{R}^K$. Agent i's valuation $u_i \in U_i$ is assumed to be his privately observed "type." A *state* is a valuation profile $(u_1, \ldots, u_N) \in U \equiv U_1 \times \ldots \times U_N \subset \mathbb{R}^{NK}$.

The objective is to implement an *allocation rule*, which is a correspondence (relation) $F : U \twoheadrightarrow K$. For each state $u \in U$, the allocation rule describes the subset $F(u) \subset K$ of "desirable" allocations. For example, we may be interested in the *efficient* allocation rule, which selects the allocations maximizing the sum of the agents' valuations (*total*

surplus):

$$F^*(u) = \arg\max_{k \in K} \sum_{i \in N} u_{ik} \ \forall u \in U.$$

Identification of efficiency with surplus-maximization is based on the ability to compensate agents with monetary transfers, and the quasilinearity of their payoffs in these transfers. Approximate efficient allocation rules will be defined in Section 5 below.

2.2 Communication

We now describe communication procedures that solve the allocation problem. It is well known that communication can be shortened by letting agents send messages sequentially rather than simultaneously. For example, an agent need not report his valuation for allocation k if previous messages have made it clear that k stands no chance of being efficient. Therefore, we must consider multi-stage communication protocols.

In the language of game theory, a general communication protocol is described with an extensive-form message game as well as each agent's strategy in this game (complete action plan contingent on his type and observed history). Instead of payoffs, the game assigns allocations to terminal nodes (and so is more properly called a "game form," or "mechanism"). The agents are assumed to follow the prescribed strategies (their incentives to do so will be discussed in Section 10). Such communication protocols are called "deterministic," because the message sent by an agent at a given information set is fully determined by his type and the preceding messages. A protocol *realizes* allocation rule F if in every state $u \in U$ it achieves a terminal node to which an allocation from $F(u)$ is assigned.

Dealing with deterministic communication protocols is quite cumbersome. Analysis can be simplified by considering what is known as "nondeterministic communication" in computer science and as the "verification scenario" in economics. Imagine an omniscient oracle who knows the true state u, and consequently knows a "desirable" allocation $k \in F(u)$, but has to prove to an ignorant outsider that k is indeed desirable. He carries out the proof by publicly announcing a message $m \in M$. Each agent i either accepts

or rejects the message, doing this on the basis of his own type u_i. The set of messages acceptable to all agents in state u is described by the message correspondence $\mu(u)$. The acceptance of message m by all agents must prove to the outsider that allocation k is desirable. This communication is called "nondeterministic" because the oracle has to "guess" an acceptable message, and there may be more than one such message in a given state.

While nondeterministic communication is patently unrealistic, we consider it for the following reasons:

1. Any deterministic communication can be represented as nondeterministic simply by letting all the messages be sent by the oracle instead of the agents, and having each agent accept the message sequence if and only all the messages sent in his stead are consistent with his strategy given his type. The oracle's message space M is thus identified with the set of the protocol's possible message sequences (terminal nodes), and the message correspondence μ is single-valued. Therefore, any lower bound on the communication requirements of nondeterministic protocols applies to deterministic protocols as a particular case.

2. A famous economic example of nondeterministic communication is the Walrasian equilibrium. The role of the oracle is played by the "Walrasian auctioneer," who announces the equilibrium prices and allocation. Each agent accepts the announcement if and only if his allocated consumption bundle maximizes his utility within the budget set given by the announced prices. The classical Welfare Theorems imply that Walrasian equilibria can be used to verify Pareto efficiency in convex exchange economies, and these results can be extended to more general price equilibria in more general allocation problems.

3. A nondeterministic protocol realizing choice correspondence F can often be used as a basis for an iterative *deterministic* protocol approximating F. At each stage of the iteration, a message $m \in M$ is announced, and each agent reports the direction in which the message should be adjusted to become "more acceptable" to

him. Such adjustment processes approximating Walrasian equilibria are known as "tâtonnement." Nondeterministic communication can then be viewed as a steady state of the deterministic adjustment process.

Formally, we define nondeterministic communication as follows:

Definition 1 *A* nondeterministic communication protocol *is a triple* $\Gamma = \langle M, \mu, h \rangle$, *where* M *is the message set,* $\mu : U \twoheadrightarrow M$ *is the message correspondence, and* $h : M \to K$ *is the outcome function, and the message correspondence* μ *has the following properties:*

- Existence: $\mu(u) \neq \emptyset$ *for all* $u \in U$,
- Privacy Preservation: $\mu(u) = \cap_i \mu_i(u_i)$ *for all* $u \in U$, *where* $\mu_i : U_i \twoheadrightarrow M$ *for all* $i \in N$.

Protocol Γ realizes *choice correspondence* $F : U \twoheadrightarrow K$ *if* $h(\mu(u)) \subset F(u)$ *for all* $u \in U$.[5]

Existence means that an acceptable message exists in each state. *Privacy Preservation* follows from the fact that each agent does not observe other agents' types when making his acceptance decision, thus the set of messages acceptable to him is a function $\mu_i(u_i)$ of his own type u_i only.[6] Finally, the definition of realization says that the acceptance of a message m by all agents proves to the outsider that $h(m) \in F(u)$.

Definition 1 has a nice interpretation in terms of geometric properties of the subsets $\mu^{-1}(m)$ of the state space U, each such subset being the event in which a given message m occurs. In this interpretation, described in Kushilevitz and Nisan (1997), Existence

[5] According to this definition, F can be realized only if it is nonempty-valued. Alternatively, we could view states in which F is empty-valued as "illegal," and allow arbitrary alternatives to be realized in such states.

[6] This is an established, if perhaps unfortunate, term in the economic literature. Privacy Preservation reflects the fact that the agents do not observe each other's types *initially*, but does not constrain the revelation of types in the course of communication.

requires that the collection $\{\mu^{-1}(m)\}_{m\in M}$ of such events cover the state space U. Privacy Preservation requires that each element of the covering be a product set $\mu_1^{-1}(m) \times \ldots \times \mu_N^{-1}(m)$—a "rectangle" in computer science parlance. Realization requires that for each rectangle $\mu^{-1}(m)$ from the covering, a single outcome $h(m)$ be "desirable" on the whole rectangle (in computer science parlance, the rectangle is "monochromatic").

2.3 The burden of discrete and continuous communication

The (nondeterministic) communication burden of an allocation rule is defined as the minimum communication burden of a (nondeterministic) protocol realizing it. The communication burden of a protocol is naturally defined as the length of the realized message sequence, i.e., the number of messages sent in the course of the protocol. Since this number may vary across states, we focus on the "worst-case" communication burden—the maximum length of the message sequence over all states (though see Section 8 below for some average-case results). For this measure to be interesting, we must require all communication to be encoded with "elementary" messages, bounding the amount of information conveyed with an elementary message.

The computer science literature on communication complexity considers discrete communication, in which the elementary messages convey a bit of information (see Kushilevitz and Nisan (1997)).[7] In particular, in the nondeterministic case, the minimal binary encoding of the oracle's message from set M takes $\log|M|$ bits. (log will stand for the binary logarithm to simplify notation).

In the case of continuous communication, agents are allowed to send real-valued elementary messages. We also want to allow finite-valued messages (say, to communicate discrete combinatorial allocations), but not count them toward the communication burden. Thus, the worst-case burden of continuous communication is defined as the maximum number of real-valued elementary messages sent in the course of the protocol. In

[7] This is merely a normalization, because an elementary message in *any* finite alphabet could be coded with a fixed number of bits.

the nondeterministic case, we can identify the communication burden with the dimension of the oracle's message space M, for which purpose we must define a topology on M.

A well-known problem in continuous communication is the possibility of "smuggling" multidimensional information in a one-dimensional message space, (e.g., using the inverse Peano function). Note, however, that with such "smuggling," a small error in the message would yield a huge error in its meaning. To avoid this, we define a metric on messages based on their meaning: the distance between messages m and m' is defined as the Hausdorff distance between the events $\mu^{-1}(m)$ and $\mu^{-1}(m')$ in which they occur.[8] The metric dimension of space M in this metric, denoted by $\dim M$, then serves as a measure of the communication burden. ($\dim M$ can be defined as the Hausdorff dimension, the box-counting dimension, or the packing index (Edgar (1990))—this would not affect the results in this paper.) With this definition, any coding of messages from M with d real numbers that prevents small errors in the code from drastically distorting the meaning (formally, the inverse of the code is Lipszhitz continuous) must have $d \geq M$ ((Edgar (1990, Exercise 6.1.9(1)). Thus, $\dim M$ is the relevant measure of communication burden if the communication must be robust to small transmission errors, due either to analog noise or to discretization ("quantization"). (See Proposition 4 below for a formal result along these lines.)

Our approach to measuring continuous communication stands in contrast to the previous literature on continuous communication, in which the topology on the message space is taken as given, and dimension smuggling is ruled out by requiring the communication protocol to satisfy a "regularity" restriction. The typical restriction, introduced by Mount and Reiter (1974) and Walker (1977), is that the message correspondence μ be "locally threaded"—i.e., have a continuous selection on a neighborhood of any point.

[8] Formally, the distance is defined as

$$\begin{aligned} \rho_M(m, m') &= \max\left\{d_M\left(\mu^{-1}(m), \mu^{-1}(m')\right), d_M\left(\mu^{-1}(m'), \mu^{-1}(m)\right)\right\}, \text{ where} \\ d_M(A, B) &= \sup_{R \in A} \inf_{R' \in B} \rho_U(R, R') \text{ for } A, B \subset U, \end{aligned}$$

with ρ_U describing the given metric on the state space U.

This restriction rules out *a priori* some important communication protocols, e.g., the communication of discrete allocations (μ cannot have a continuous selection in a neighborhood in which the allocation switches).

3 Efficient Communication and Supporting Prices

One particular way to verify the efficiency of an allocation is by announcing supporting prices:

Definition 2 *A pair* $(p,k) \in \mathbb{R}^{NK} \times K$*, where* $k \in K$ *is the proposed allocation and* $p \in \mathbb{R}^{NK}$ *is a list of personalized allocation prices, is a* price equilibrium *in state* $u \in U$ *if it satisfies the following inequalities:*

$$u_{ik} - p_{ik} \geq u_{ik'} - p_{ik'} \text{ for all } i \in N,\ k' \in K. \tag{1}$$

$$\sum_{i \in N} p_{ik} \geq \sum_{i \in N} p_{ik'} \text{ for all } k' \in K. \tag{2}$$

(1) says that the proposed allocation maximizes each agent's utility net of the announced prices. (2) can be interpreted as requiring that the proposed allocation maximize the designer's revenue given the announced prices. Denote the set of price equilibria in state u by $E(u)$, and let $E : U \twoheadrightarrow \mathbb{R}^{NK} \times K$ denote the price equilibrium correspondence.

A price equilibrium with $N = 2$ agents is illustrated in Figure 1. Without loss of generality we normalize both agents' utilities and prices for the equilibrium allocation k to zero, and graph agent 1's valuations and prices for all allocations in the downward direction, and those of agent 2 in the upward direction. (The economic interpretation of Figure 1 is as an "Edgeworth box" whose vertical dimension represents the split of money between the two agents, and whose horizontal dimension represents the allocations. In this interpretation, u_1 and u_2 depict the agents' respective indifference curves passing through the equilibrium point, and p_1 and p_2 depict the boundaries of their respective budget sets.) (2) says that the curve representing p_1 must lie above that representing p_2, and (1) that the curve representing u_1 is above that for p_1 and the curve representing u_2 is below that for p_2.

This general notion of price equilibrium was first suggested by Mas-Colell (1980), and later used by Bikhchandani and Mamer (1997) and Bikhchandani and Ostroy (2002) for combinatorial allocation problems. These papers have made the observation that all price equilibrium allocations are efficient, often referred to as the *First Welfare Theorem*. In Figure 1, efficiency simply means that u_1 must lie above u_2.

The First Welfare Theorem allows to verify efficiency with a *price protocol*, whose message space is $M \subset \mathbb{R}^{NK} \times K$, whose message correspondence is $\mu(u) = E(u) \cap M$, and whose outcome function is $h(p,k) = k$. Note that we allow many different price protocols described by different message spaces, i.e., different feasible price-allocation pairs. Any price protocol satisfies *Privacy Preservation* by construction, since each agent verifies his inequality in (1) using his own type only, and any agent can verify (2). If we find a message space M large enough to satisfy *Existence* (i.e., that a price equilibrium from M exists in all states), the size of M provides an upper bound on the burden of efficient communication. (Namely, it is $\log|M|$ bits for discrete communication, and $\dim M$ for continuous communication).

One might think that among all possible efficient communication protocols, there would be some that are distinct from, and perhaps much simpler than, price protocols. However, we establish that this is not the case— any efficient communication protocol must reveal supporting equilibrium prices:[9]

Proposition 1 *Communication protocol $\Gamma = \langle M, \mu, h \rangle$ realizes the efficient allocation rule F^* if and only if there exists an assignment $p : M \to \mathbb{R}^{NK}$ of prices to messages such that protocol $\langle M, \mu, \langle p, h \rangle \rangle$ realizes the price equilibrium correspondence E.*

Proof. The "if" statement obtains by adding up all inequalities (1) and (2). For the "only if" statement, suppose protocol $\langle M, \mu, h \rangle$ realizes F^*. For each $m \in M$, let $\hat{k} = h(m)$, and let $p_{ik} = \sup_{u_i \in \mu_i^{-1}(m)} [u_{ik} - u_{i\hat{k}}]$ for all $i \in N$, $k \in K$. By construction,

[9] A similar result is established by Parkes (2002), but only for a restricted communication language.

$\left(p, \hat{k}\right)$ satisfies (1) for each $u \in \mu^{-1}(m)$. At the same time, for all $k \in K$ we can write

$$\sum_i [p_{ik} - p_{i\hat{k}}] = \sum_i \sup_{u_i \in \mu_i^{-1}(m)} [u_{ik} - u_{i\hat{k}}] = \sup_{u \in \mu^{-1}(m)} \sum_i [u_{ik} - u_{i\hat{k}}] \leq 0,$$

using *Privacy Preservation* and the fact that $\hat{k} \in F^*(u)$ for each $u \in \mu^{-1}(m)$. Thus, $\left(p, \hat{k}\right)$ also satisfies (2). Therefore, $\left(p, \hat{k}\right)$ is a price equilibrium in every state $u \in \mu^{-1}(m)$. ■

The "if" statement of the Proposition is the First Welfare Theorem. The "only if" statement can be thought of as a strengthening of the traditional Second Welfare Theorem, which says only that for any efficient allocation we can construct supporting prices *given full information about the economy*. For general price equilibria, the construction is trivial: for example, we can simply take prices $p^i = u^i$ for all agents i. The Second Welfare Theorem is not a very useful result, because with full information, an efficient allocation can be implemented directly, without using prices. In contrast, Proposition 1 says not only that supporting prices exist, but also that they must be revealed by *any* efficient communication, not just by full revelation.

The proof of the "only if" part of Proposition 1 for two agents is illustrated in Figure 2, which depicts the valuations in the same way as Figure 1. Suppose that a message m has verified that allocation k is efficient. Consider the two agents' valuations that are consistent with message m. Due to *Privacy Preservation*, k must be efficient in any state in which agent 1's valuation is consistent with m and agent 2's valuation is consistent with m. Graphically, this means that any valuation curve of agent 1 consistent with m (in Figure 2, u_1 and u_1') must lie above any valuation curve of agent 2 consistent with m (in Figure 2, u_2 and u_2'). Therefore, letting agent 1's price curve p_1 be the lower envelope of his valuation curves consistent with m, and letting agent 2's price curve p_2 be the *upper* envelope of his valuation curves consistent with m, p_1 will lie above p_2, thus the prices will satisfy condition (2). Also, by construction, the prices will satisfy condition (1) in all states consistent with m (states (u_1, u_2), (u_1', u_2), (u_1, u_2'), and (u_1', u_2') in Figure 1) Thus, on the basis of a message m verifying k we have constructed a price equilibrium supporting k in all states in which m occurs.

Proposition 1 implies that the nondeterministic communication burden of the efficient correspondence F^* is *exactly* the minimum size of a price-allocation space M ensuring the existence of an equilibrium from $M \subset \mathbb{R}^{NK} \times K$. While finding such a minimal space is in general a hard problem, below we derive some upper and lower bounds for it in specific cases (and in some of these cases the two bounds match).

4 A Lower Bound on Efficient Communication

Note that two utility functions that differ by a constant describe the same preferences. In counting the states of the world, we will need to ensure that we do not count two different states describing the same preferences. Similarly, in measuring the metric dimension of (subsets of) the state space, we need to ensure that the dimensionality of utility functions coincides with the dimensionality of preferences. For these purposes, we assume that any two preferences that are "close" to each other are described by utility functions that are close to each other:

Definition 3 *The state space U is* normalized *if there exists $C > 0$ such that for all $u, u' \in U$,*[10]

$$\|u' - u\| \leq C \sup_{i \in N,\ j,k \in K} \left| \left[u'_{ij} - u_{ij}\right] - \left[u'_{ik} - u_{ik}\right] \right|.$$

In particular, a normalized state space U cannot contain two distinct states between which each agent's utility differs only by a constant. For example, normalization holds (with $C = 1$) when each agent is always assigned zero utility to one of the allocations.

Now we can show that the subset of the state space on which all allocations are equally efficient constitutes a "fooling set" for the price equilibrium correspondence. Namely, letting

$$U^* = \{u \in U : F^*(u) = K\},$$

we show that two distinct states from U^* cannot have the same supporting prices:

[10] For definiteness, we let $\|\cdot\|$ represent the sup-norm on $\mathbb{R}^{NK}$, though in a finite-dimensional space U all norms are known to be equivalent .

Lemma 1 *In a normalized quasilinear problem, for any $u, u' \in U^*$ with $u \neq u'$, $E(u) \cap E(u') = \emptyset$.*

Proof. Suppose in negation that $\left(p, \hat{k}\right) \in E(u) \cap E(u')$. $u \in U^*$ means that the total surplus $\sum_i u_{ik}$ does not depend on $k \in K$, and therefore (2) can be written in state u as

$$\sum_i [u_{i\hat{k}} - p_{i\hat{k}}] \leq \sum_i [u_{ik} - p_{ik}] \text{ for all } k \in K.$$

On the other hand, (1) means that

$$u_{i\hat{k}} - p_{i\hat{k}} \leq u_{ik} - p_{ik} \text{ for all } i \in N,\ k \in K.$$

Comparing the two displays, we see that for each i, $u_{ik} - p_{ik}$ does not depend on $k \in K$. Since by the same argument the same is true for $u'_{ik} - p_{ik}$, we see that $u'_{ik} - u_{ik}$ does not depend on k for all i. By normalization, this implies that $u = u'$ —a contradiction. ■

The proof of Lemma 1 for the case $N = 2$ can be illustrated graphically in Figure 1, by observing that in any state from U^*, the two agents' valuation curves coincide, and the price curves are squeezed in-between. Therefore, both agents' prices must coincide with their utilities (up to a constant). Different states from U^* will give rise to different indifference curves and will therefore pin down different price curves.

We can now use Proposition 1 to conclude that any efficient communication protocol must yield distinct messages in distinct states from U^*. This implies that the protocol uses at least $|U^*|$ distinct messages. A similar statement can be made for continuous communication, though extra care should be taken to rule out "dimension smuggling":

Proposition 2 *In a normalized quasilinear problem, any efficient protocol transmits at least* $\log |U^*|$ *bits, and the dimension of its message space is at least* $\dim U^*$.

Proof. Lemma 1 and Proposition 1 together imply that the restriction of the message correspondence μ to U^* has an injective selection $\sigma : U^* \to \mu(U^*)$. This implies the first statement of the proposition. For the second statement, we will also show that σ^{-1} is Lipschitz continuous. For this purpose, take any $m, m' \in \mu(U^*)$, and let $u = \sigma^{-1}(m)$,

and $u' = \sigma^{-1}(m')$. By the construction of the metric on messages (see footnote 8), there exists $u'' \in \mu^{-1}(m')$ such that $\|u'' - u\| \leq \rho(m', m)$. Since $u' \in U^*$, as shown in the proof of Lemma 1, any equilibrium price vector p in state u' coincides with u' up to agent-dependent constants. By Proposition 1, price vector u' must also support allocation $h(m') = \hat{k}$ in state u''. (1) then implies that for all $i \in N$, $k \in K$,

$$u''_{ik} - u''_{i\hat{k}} \leq u'_{ik} - u'_{i\hat{k}} = u'_{-i,\hat{k}} - u'_{-i,k} \leq u''_{-i,\hat{k}} - u''_{-i,k}$$

(where $u_{-i,k} = \sum_{j \neq i} u_{jk}$). Subtracting $u_{ik} - u_{i\hat{k}} = u_{-i,\hat{k}} - u_{-i,k}$, and taking the absolute value implies:

$$\begin{aligned} \left|(u'_{ik} - u_{ik}) - (u'_{i\hat{k}} - u_{i\hat{k}})\right| &\leq \max\left\{\left|(u''_{ik} - u_{ik}) - (u''_{i\hat{k}} - u_{i\hat{k}})\right|, \left|(u''_{-i,k} - u_{-i,k}) - (u''_{-i,\hat{k}} - u_{-i,\hat{k}})\right|\right\} \\ &\leq 2(N-1)\|u'' - u\|. \end{aligned}$$

Now, using normalization and the triangle inequality,

$$\begin{aligned} \|u' - u\| &\leq C \sup_{i \in N,\ j,k \in K} \left|\left(u'_{ij} - u_{ij}\right) - (u'_{ik} - u_{ik})\right| \\ &\leq 2C \sup_{i \in N,\ k \in K} \left|(u'_{ik} - u_{ik}) - (u'_{i\hat{k}} - u_{i\hat{k}})\right| \\ &\leq 4C(N-1)\|u'' - u\| \leq 4C(N-1)\rho(m', m). \end{aligned}$$

Thus, σ^{-1} is Lipschitz continuous, which implies that $\dim M \geq \dim \sigma(U^*) \geq \dim U^*$ (see Edgar (1990, Exercise 6.1.9(1)). ■

The simplest application of Proposition 2 is to the case $N = 2$:

Corollary 1 *Suppose that in a normalized quasilinear problem with $N = 2$, for each $u_1 \in U_1$ there exists a "dual utility" $u_2 \in U_2$ such that $u_{1k} + u_{2k}$ does not depend on k. Then any efficient protocol transmits at least $\log|U_1|$ bits and has the message space dimension of at least $\dim U_1$.*

Corollary 1 will be a workhorse for the subsequent results on the combinatorial allocation problem. We will apply it also to $N > 2$ agents by letting agents $i > 2$ have constant utilities over all allocations k. Also, Corollary 1 can be applied to some cases in

which the two agents can have dual utilities only on some subset $\tilde{K} \subset K$ of allocations, by restricting attention to a subclass of utilities for which allocations from $K \backslash \tilde{K}$ are never efficient (see Subsection 7.1).

5 Approximation and Discretized Problems

One may hope that approximate efficiency could be achieved with less communication than exact efficiency. In this section we discuss how to analyze the communication burden of approximate efficiency. To be consistent with the computer science literature on approximation (see, e.g., Vazirani (2001)), we use an approximation measure that is invariant to the units of measurement (see, e.g., Vazirani (2001)). Namely, defining the maximum surplus available in state u by

$$S(u) = \max_{k \in K} \sum_i u_{ik},$$

we define the choice correspondence F_r^* realizing approximation ratio $r \in [0,1]$ as follows:[11]

$$F_r^*(u) = \left\{ k \in K : \sum_i u_{ik} \geq rS(u) \right\}.$$

By construction, $F_1^* = F^*$ (the exactly efficient correspondence), and the set $F_r^*(u)$ is nonincreasing in r for all u.

[11] This is a "worst-case" definition, since it requires uniform approximation across all states. The weaker requirement of "average-case" approximation given some probability distribution over states is considered in Section 8 below.

For the study of approximation, we need to assume[12]

$$\overline{S} \equiv \sup_{u \in U} S(u) < \infty, \underline{S} \equiv \inf_{u \in U: S(u) \neq 0} S(u) > 0. \tag{3}$$

The communication burden of F_r^* can be evaluated by examining the communication burden of the exactly efficient correspondence F^* in a discretized problem. In computer science, discrete problems are usually considered in their own right (the inputs are restricted to be integers), but we will use them as a stepping stone for the analysis of approximation in the continuous problem.

For each state $u \in U$, let u^δ denote the state in which all utilities are rounded off to multiples of $\delta > 0$. Define the *upper δ-discretized problem* as the problem with the discrete state space $U^\delta = \left\{u^\delta : u \in U\right\}$. Suppose that we have a protocol Γ realizing exact efficiency for problem U^δ. We can then ask the agents to round off their utilities to multiples of δ and follow protocol Γ. Since the sum of rounded-off utilities for every allocation is within $N\delta/2$ from the true surplus at this allocation, the maximization of this sum results in a surplus loss of at most $N\delta$. Since the maximum available surplus is bounded below by $\underline{S}$, we realize approximation ratio $1 - N\delta/\underline{S}$, using as much communication as in Γ. In particular, full revelation of valuations rounded off with a sufficiently fine precision achieves an approximation ratio arbitrarily close to 1.

Examination of a discretized problem also allows to bound the communication burden of approximation *from below*. For this purpose, define the *lower δ-discretized problem* as the problem with the discrete state space $U^\delta \cap U$. This state space consists of those valuations from U that are multiples of δ. (In most applications considered in this paper,

[12]Both inequalities are needed to ensure that approximation can be achieved with finite communication. For example, consider the problem of allocating one object between two agents whose valuations lie in $[0,1]$, and so $\underline{S} = 0$. Pick $r \in (0,1)$, and consider the restricted problem in which both agents' valuations lie in the set $\{r^n\}_{n=0}^\infty \subset [0,1]$. In this restricted problem, realizing an approximation ratio higher than r is equivalent to exact efficiency, and Corollary 1 implies that this requires a countable message space. Since arbitrary $r \in (0,1)$ can be chosen, this implies that no positive approximation can be achieved with finite communication. The same conclusion is reached when the agents' valuations lie in $[1,\infty]$ (and so $\overline{S} = \infty$), by considering the restricted problem $\{r^{-n}\}_{n=0}^\infty$.

$U^\delta \subset U$, hence the lower and upper discretized problems coincide, in which case we will simply call them the discretized problem.) Since any misallocation in problem $U^\delta \cap U$ loses at least surplus δ, and the maximum available surplus is bounded above by $\bar{S}$, realizing an approximation ratio higher than $1 - \delta/\bar{S}$ in problem $U^\delta \cap U$ is equivalent to realizing exact efficiency. Realizing the same approximation ratio in problem U requires at least as much communication. This observation will allow us to bound below the communication burden of approximating efficiency in problem U by applying Corollary 1 to the upper discretized problem. To summarize:

Proposition 3 *In a quasilinear problem in which (3) holds, (i) realizing approximation ratio $1 - \delta N/\underline{S}$ does not require more communication than realizing exact efficiency in the upper discretized problem U^δ, and (ii) realizing an approximation ratio higher than $1-\delta/\bar{S}$ requires at least as much communication as realizing exact efficiency in the lower discretized problem $U^\delta \cap U$.*

We want to characterize the dependence of the communication burden on the desired approximation, as well as on the parameters of the problem. For this purpose, we use the following three concepts, listed from weaker to stronger, which are standard in the computer science literature on approximation algorithms (see, e.g., Vazirani (2001)):

- A *Polynomial Approximation Scheme* (PAS) in some parameters is a protocol that for any given $\varepsilon > 0$ realizes approximation ratio $1 - \varepsilon$ using a number of bits that is polynomial in the parameters.

- A *Fully Polynomial Approximation Scheme* (FPAS) in some parameters is a protocol that for any $\varepsilon > 0$ realizes approximation ratio $1 - \varepsilon$ using a number of bits that is polynomial in ε^{-1} and the parameters.

- A *Truly Polynomial Approximation Scheme* (TPAS) in some parameters is a protocol that for any $\varepsilon > 0$ realizes approximation ratio $1 - \varepsilon$ using a number of bits that is polynomial in $\log \varepsilon^{-1}$ and the parameters.

PAS achieves an arbitrarily close approximation with polynomial communication, but does not stipulate how the communication burden depends on the approximation error. In contrast, in FPAS and TPAS, the error must shrink sufficiently fast with the number of bits transmitted. An economic example of FPAS is an ascending-bid auction for one unit with a minimum bid increment ε. Suppose that the agents' valuations for the unit lie in $[1, 2]$. At each price level starting from $p = 1$, the auction asks each agent to send one bit—"in" or "out." If at least one agent sends "in," the price is incremented by ε. The auction stops when all agents send "out," assigning the object to (one of) the agent(s) who sent "in" in the previous stage. Suppose that the agents behave "truthfully"—send "in" if and only if their valuations exceed the current price. Then the auction is exactly efficient for the discretized problem U^ε, and therefore, by Proposition 3(i), it realizes approximation ratio $1 - \varepsilon$. Since the maximum number of price increments is ε^{-1}, the auction's worst-case complexity is $N\varepsilon^{-1}$.

TPAS requires a much faster approximation than FPAS—the error must now shrink exponentially with the number of bits transmitted. An economic example of TPAS is a sealed-bid auction of a single unit, in which the agents submit their valuations rounded off to a multiple of ε. Suppose the agents' valuations for the unit lie in $[1, 2]$. Since the auction is exactly efficient for the upper discretized problem U^ε, by Proposition 3(i), it realizes approximation ratio $1 - \varepsilon$. Since it takes $\log \varepsilon^{-1}$ bits to transmit a valuation rounded off to a multiple of ε, the total number of bits transmitted is $N \log \varepsilon^{-1}$.

Note that our TPAS example was obtained by taking a fully efficient continuous protocol and asking the agents to round off their messages. This technique can be generalized: given a d-dimensional continuous protocol realizing approximation ratio r, rounding off the messages yields a TPAS to approximation ratio r that is linear in d. Intuitively, a message round-off error that is small in the Hausdorff metric defined in footnote 8 yields an allocation that is desirable for some state that is not too far from the true state, and therefore a small efficiency loss. Formally, we have

Proposition 4 *In a quasilinear problem satisfying (3), if there exists a protocol realizing*

approximation ratio r with a message space of upper box dimension $d > 0$,[13] *then for any $\varepsilon > 0$ there exists a protocol realizing approximation ratio $r - \varepsilon$ using $C(\varepsilon)$ bits, with $C(\varepsilon) \sim d \log \varepsilon^{-1}$ as $\varepsilon \to 0$.*

Proof. Suppose that $\Gamma = \langle M, \mu, h \rangle$ is an efficient protocol. By the definition of upper box dimension, for each $\delta > 0$ there exists a covering of M by balls of radius δ centered at points $M_\delta \subset M$, with $\log |M_\delta| \sim d \log \delta^{-1}$. For each $m \in M$, we can choose a point $r_\delta(m) \in M_\delta$ such that $\rho(m, r_\delta(m)) \leq \delta$. Let $\mu_\delta(u) = r_\delta(\mu(u))$.

Consider the protocol $\Gamma_\delta = \langle M_\delta, \mu_\delta, h \rangle$. Note that Γ_δ inherits Existence and Privacy Preservation from Γ. Take any state $u \in U$ and any $m_\delta \in \mu_\delta(u)$. By construction, $m_\delta = r_\delta(m)$ for some $m \in \mu(u)$. By the construction of the Hausdorff metric ρ on messages (see footnote 8), there exists $u' \in \mu^{-1}(m_\delta)$ such that $\|u' - u\| \leq \rho(m, m_\delta) \leq \delta$. This implies that (i) $S(u) \leq S(u') + \delta$, and (ii) $\sum_i u_i(h(m_\delta)) \geq \sum_i u'_i(h(m_\delta)) - \delta$. Thus, we can write

$$\frac{\sum_i u_i(h(m_\delta))}{S(u)} \geq \frac{\sum_i u'_i(h(m_\delta)) - \delta}{S(u') + \delta} \geq \frac{\sum_i u'_i(h(m_\delta)) - 2\delta}{S(u')} \geq r - 2\delta/\underline{S},$$

where the last inequality uses the fact that Γ realizes approximation ratio r, hence $h(m_\delta) \in F_r^*(u')$. Therefore, taking $\delta(\varepsilon) = \varepsilon \underline{S}/2$ ensures that $\Gamma_{\delta(\varepsilon)}$ realizes approximation ratio $r - \varepsilon$. The number of bits communicated by $\Gamma_{\delta(\varepsilon)}$ is

$$\log \left| M_{\delta(\varepsilon)} \right| \sim d \log \delta(\varepsilon)^{-1} = d \log \varepsilon^{-1} + d \log(2/\underline{S}) \sim d \log \varepsilon^{-1} \text{ as } \varepsilon \to 0. \blacksquare$$

In particular, the Proposition implies that if we discretize two efficient continuous protocols with message spaces of dimension d_1 and d_2 respectively to guarantee approximation error ε, then the worst-case complexities of the two discretized protocols are related as d_1/d_2 asymptotically as $\varepsilon \to 0$.[14] Thus, the dimension of the continuous message space based on the Hausdorff metric over messages is indeed a relevant measure of

[13] The upper box dimension coincides with the Hausdorff dimension for most "well-behaved" sets, but could exceed it for some sets (see Edgar (1990)).

[14] Conversely, if a given large number of bits C is to be transmitted in both cases, then the ratio of the approximation errors of the two protocols will grow exponentially with C when $d_2 < d_1$. This is consistent with the findings of Hurwicz and Marschak (2003a,b).

communication even if real-life communication is discrete. The result also implies that if we have continuous communication that is polynomial in some parameters and realizes approximation ratio r, then the discretized protocol is a TPAS for r in the same parameters. In such a case we will simply say that the discretized protocol is a "polynomial protocol realizing approximation ratio r."

6 The Combinatorial Allocation Problem

We now specialize to the *Combinatorial Allocation (CA) problem*, in which the allocation is that of L items among the N agents. Formally, the allocation space is $K = N^L$, and $k(l)$ denotes the agent holding item $l \in L$ in allocation $k \in K$.

We impose several standard restrictions on valuations (which can only reduce the communication burden):

- **No Externalities (NE):** For each i and each $u_i \in U_i$, there exists $v_i : 2^L \to \mathbb{R}$ such that $u_{ik} = v_i(k^{-1}(i))$ for each $k \in K$.

In words, each agent i's utility is a function v_i of the bundle $k^{-1}(i)$ allocated to him. We will call v_i the agent's *valuation*, and let $V_i \subset \mathbb{R}^{2^L}$ denote the class of his possible valuations. The state space can then be represented by $V = V_1 \times \ldots \times V_N$.

Each $v_i \in V_i$ is also assumed to satisfy the following restrictions:

- **Normalization (N):** $v_i(\emptyset) = 0$.
- **Monotonicity (M):** $v_i(S)$ is nondecreasing in $S \subset L$.
- **Boundedness (B):** Either $v_i(L) = 0$ or $v_i(L) \in [\gamma, 1]$.

(N) is without loss of generality, and it serves to rule out distinct valuations that differ only by a constant, thus describing the same preferences. It also ensures the Normalization assumption above. (M) and (B) are not needed for the analysis of exact efficiency, but will be used in the analysis of approximation (in particular, (B) ensures

condition (3) above). Given that the analysis is invariant to scale, the choice of the upper bound as 1 is arbitrary. The lower bound γ can be viewed as a parameter of the problem along with N and L, but its value will be irrelevant for most results. Let $V_{\text{gen}} \subset \mathbb{R}^{2^L}$ denote the class of all valuations satisfying (NE), (N), (M), and (B). (We will consider additional restrictions on valuations in Section 7.)

To apply Corollary 1, for each valuation $v \in V_{\text{gen}}$ define the *dual* valuation $\tilde{v} \in V_{\text{gen}}$ by

$$\tilde{v}(S) = v(L) - v(L \backslash S) \text{ for all } S \subset L. \tag{4}$$

When $N = 2$ and the two agents' valuations are v and $\tilde{v}$, all allocations have the same surplus $v(L)$. Thus, we can use Corollary 1 to obtain

Proposition 5 *In the combinatorial allocation problem with general valuations, the dimension of the message space in any efficient protocol is at least* $\dim V_{gen} = 2^L - 1$.

Therefore, the communication burden of efficiency is at least as large as a full description of one agent's valuations. Recall from the discussion in Section 3 that this lower bound is essentially tight for $N = 2$. A more general upper bound is given by $(N-1)\left(2^L - 1\right)$ numbers (all agents but one fully reveal their valuations), but the exact communication burden for $N > 2$ remains an open problem.[15]

To bound below the communication burden of approximation, we apply the same logic to the discretized problem $V^{\delta}_{\text{gen}} \subset V_{\text{gen}}$ with $\delta = 1$ (i.e., the problem in which the agents' valuations for all bundles are either 0 or 1). Corollary 1 implies that any efficient protocol for the discretized problem transmits at least $\log \left|V^1_{\text{gen}}\right|$ bits. Only counting those

[15] To see that the upper bound is not tight either, consider the case where $N > L$. In this case, in any allocation at most L agents receive non-empty bundles, therefore it cannot be efficient to allocate a bundle to an agent who does not hold one of the top L valuations for it. Thus, efficiency can be verified by announcing only the L highest valuations for each bundle (and the agents holding them), and having each agent i accept the communication if for each bundle $S \subset L$, either his valuation for S is announced correctly, or it does not exceed any of the L announced valuations. Therefore, when $N > L$, efficiency can be verified using only $L\left(2^L - 1\right)$ real numbers. We are obliged to Moshe Babayoff for bringing this to our attention.

valuations in V^1_{gen} that have $v(S) = 0$ for $|S| < L/2$ and $v(S) = 1$ for $|S| > L/2$, we see that $\log \left|V^1_{\text{gen}}\right| \geq \binom{L}{L/2}$.[16] Thus, already with $N = 2$, by Proposition 3(ii) we obtain a lower bound on the communication burden of realizing a higher approximation ratio than $1 - \delta/N = 1/2$:

Proposition 6 *In the combinatorial allocation problem with general valuations, any protocol realizing an approximation ratio higher than* $1/2$ *communicates at least* $\log \left|V^1_{gen}\right| \geq \binom{L}{L/2}$ *bits.*

Observe that approximation ratio $1/N$ can be realized by auctioning off all objects as a bundle to the highest bidder. (Indeed, the bundled auction realizes surplus $\max_i v_i(L)$, while no individual agent can have a higher utility than that at any allocation.) Thus, Proposition 6 means that for $N = 2$, any improvement upon the bundled auction still requires very extensive communication, which still grows exponentially with L.[17] In fact, we can prove a similar statement for $N > 2$, though using a different proving technique:

Proposition 7 *In the combinatorial allocation problem with general valuations, realizing an approximation ratio higher than* $1/N$ *requires communicating at least* $\ln 2 \cdot \exp\left\{L/\left(2N^2\right) - 2\ln N\right\}$ *bits.*

Proof. In the Appendix we consider the following *set packing problem*: Each of the N agents holds a collection of subsets of L, and the goal is to approximate the maximum *packing number*—the number of subsets in the union of their collections that are packed together, i.e., are pairwise disjoint. The set packing problem is reduced to the combinatorial allocation problem by letting, for each agent i, $v_i(S)$ be the maximum number of subsets in his collection that can be packed into S. We prove a lower bound on the

[16] Note that $\left|V^1_{gen}\right|$ is the number of monotone boolean functions of L boolean variables. The problem of counting these functions is known as "Dedekind's problem," which is unsolved, though its asymptotic behavior is obtained by Korshunov (1981).

[17] Indeed, by Stirling's formula, the communication burden $\binom{L}{L/2} \sim \sqrt{2/(\pi L)} \cdot 2^L$ as $L \to \infty$.

communication complexity of any protocol realizing an approximation ratio higher than $1/N$ for the maximum packing number, which applies even when each agent i's individual packing number (corresponding to $v_i(L)$) is restricted to be at most 1. This lower bound is given in the Proposition. ■

This result should be contrasted with the findings of Lehmann et al. (1999) and Holzman et al. (2001), who suggest "simple" protocols improving upon the bundled auction. For example, Holzman et al. (2001) note that auctioning off the objects in two equal-sized bundles achieves approximation ratio $r(L) = 2/L$ for any N, thus improving upon the bundled auction when $N > L/2$ (splitting L into more bundles allows further improvement). Lehmann et al. (1999) propose a polynomial approximation algorithm that can be adapted to the following protocol: At each stage of the protocol, each agent i who is not yet allocated any items announces a subset S_i of yet unallocated items that maximizes the ratio $v_i(S_i)/\sqrt{|S_i|}$, along with the maximum ratio itself. The agent who announces the highest ratio receives the requested bundle and quits. In the course of this protocol, agents announce $N(N+1)/2$ valuations and bundles. Lehmann et al. (1999) show that this algorithm realizes approximation ratio $r(L) = 1/\sqrt{L}$, which is higher than that realized by the bundled auction when $N > \sqrt{L}$.

Observe that these improvements over the bundled auction do no contradict Proposition 7. Intuitively, the proposition implies that in large problems in which the number N of agents is "substantially smaller" than the number L of items (e.g., smaller than $L^{1/2-\varepsilon}$), "simple" protocols (e.g., polynomial in L) cannot improve over the bundled auction. When N is either comparable with or larger than L, simple protocols *can* improve over bundled auctions, though both bundled auctions and all other simple protocols realize a vanishing share of the available surplus as $N, L \to \infty$.

7 Restricted Valuations

7.1 Submodular Valuations

Here each agent's valuation space $V_i = V_{\text{sm}}$ is the set of all valuations $v \in V_{\text{gen}}$ that satisfy

$$v(S \cup T) + v(S \cap T) \leq v(S) + v(T) \text{ for all } S, T \subset L.$$

An equivalent definition of submodularity is that the marginal benefit of each item $l \in L$, $v(S \cup l) - v(S)$, is nonincreasing in $S \subset L$.

Corollary 1 cannot be applied to this case directly, since the dual (4) of a submodular valuation is typically not submodular (unless both are additive—see Subsection 7.3 below). We get around this problem by defining duality in such a way that the surplus is constant only on the allocations involving even splits of objects:

$$\tilde{K} = \left\{k \in K : \left|k^{-1}(1)\right| = \left|k^{-1}(2)\right| = L/2\right\}.$$

Namely, consider the set $\tilde{V}$ of valuations $v \in \mathbb{R}^{2^L}$ satisfying

- $v(S) = 2|S|/L$ for $|S| < L/2$,
- $v(S) = 1$ for $|S| > L/2$,
- $v(S) \in [1 - 1/L, 1]$ for $|S| = L/2$,
- $\frac{1}{\left|\tilde{K}\right|} \sum_{S \subset L : |S| = L/2} v(S) = 1 - 1/(2L)$.

One can easily verify that $\tilde{V} \subset V_{\text{sm}}$.

Note that in any state $(v^1, v^2) \in \tilde{V} \times \tilde{V}$, all efficient allocations lie in $\tilde{K}$. (Indeed, the last bullet implies that the average surplus of all allocations from $\tilde{K}$ is $2 - 1/L$, while any other allocation has a surplus of at most $2 - 2/L$, and thus is dominated by at least one allocation from $\tilde{K}$.) The last bullet above ensures that thus constructed valuation class with allocations restricted to $\tilde{K}$ is normalized.

For each $v \in \tilde{V}$, define its "quasi-dual" $\widehat{v} \in \tilde{V}$ as follows:

- $\widehat{v}(S) = 2 - 1/L - v(L \backslash S)$ for $|S| = L/2$,
- $\widehat{v}(S) = 2|S|/L$ for $|S| < L/2$.
- $\widehat{v}(S) = 1$ for $|S| > L/2$,

By construction, the set of efficient allocations in any state $(v, \widehat{v}) \in \tilde{V} \times \tilde{V}$ is exactly $\tilde{K}$. Thus, we can apply Corollary 1 with the two agents' valuations restricted to $\tilde{V}$ and the allocations restricted to $\tilde{K}$, which yields

Proposition 8 *In the combinatorial allocation problem with submodular valuations, the dimension of the message space in any efficient protocol is at least* $\dim \tilde{V} = \left|\tilde{K}\right| - 1 = \binom{L}{L/2} - 1$.

Consider now the discretized problem $\tilde{V}^{\delta} \subset \tilde{V}$ with $\delta = 1/L$ and $N = 2$. Applying Corollary 1 with the two agents' valuations restricted to $\tilde{V}^{\delta}$ and the outcomes restricted to $\tilde{K}$ implies that the number of bits communicated by an efficient protocol for the discretized problem is at least $\log \left|\tilde{V}^{\delta}\right| = \left|\tilde{K}\right| - 1$. Proposition 3(ii) then implies

Proposition 9 *With submodular valuations, realizing an approximation ratio higher than* $1 - 1/(2L)$ *requires communicating at least* $\binom{L}{L/2} - 1$ *bits.*

Note that if we had a FPAS in L, then it could be used to realize approximation ratio $1 - 1/(2L)$ using polynomial communication in L, contradicting Proposition 9. Therefore, we have

Corollary 2 *With submodular valuations, FPAS in L is impossible.*

This result implies, for example, that an ascending-bid auction with L per-item prices and bid increment ε cannot approximate efficiency within ε, because the auction's worst-case complexity would be $NL\varepsilon^{-1}$, and so it would be a FPAS. Yet, we have been unable to rule out PAS – i.e., achieving any approximation ratio with polynomial communication in N, L. We do know from Lehmann et al. (2001) that approximation ratio $1/2$ is realized

by the (deterministic) "greedy" protocol that allocates the objects in a fixed order to the agents who announce the highest current marginal benefit for them, and so communicates only NL numbers.

7.2 Homogeneous Valuations

Here each agent's valuation space $V_i = V_{\text{h}}$ is the set of valuations $v \in V_{\text{gen}}$ that satisfy $v(S) = \phi(|S|)$ for all $S \subset L$, where $\phi : \{0, \ldots, L\} \to \mathbb{R}$. That is, agents care only about the number of items they receive.

Since the dual (4) of a homogeneous valuation is homogeneous, Corollary 1 implies

Proposition 10 *In the combinatorial allocation problem with homogeneous valuations, the dimension of the message space in any efficient protocol is at least* $\dim V_h = L$.

Now consider the discretized problem $V_{\text{h}}^{\delta} \subset V_{\text{h}}$, with $\delta = 1/R$ for some R. By Corollary 1, exact efficiency in this problem requires communicating at least $\log \left|V_{\text{h}}^{\delta}\right|$ bits. Note that $\left|V_{\text{h}}^{\delta}\right|$ is the number of monotone functions $\phi : \{1, \ldots, L\} \to \{0, \ldots, R\}$ such that $\phi(L) \geq \gamma R$. If we fix $\phi(L) = R$ for simplicity, then the number of such functions is exactly the number of ways that R indistinguishable balls (corresponding to the function's unit jumps) can be partitioned into L urns (corresponding to the jump points). This number is $\binom{R+L-1}{L-1}$.[18] Therefore, exact efficiency in problem V_{h}^{δ} requires communicating at least $\log \binom{\delta^{-1}+L-1}{L-1}$ bits. By Proposition 3(ii), at least as much communication is needed to realize an approximation ratio higher than $1 - \delta/2$ when $N = 2$.

On the other hand, the full revelation protocol in the discretized problem V_{h}^{δ} uses at most $N\delta^{-1} \log(L+1)$ bits. Indeed, each agent needs only to communicate δ^{-1} jump points in $\{0, \ldots, L\}$, each of which is communicated with $\log(L+1)$ bits. By Proposition 3(i), this communication realizes approximation ratio $1 - \delta/\gamma$ in the continuous

[18] This is proven by putting all the R balls in a row with $L-1$ dividers, and letting urn $l = 1, .., L$ contain the balls lying between dividers $l-1$ and l. Thus, the different allocations of balls into urns correspond to the different positions that the $L-1$ dividers can occupy in a row of $R+L-1$ objects.

problem (note that the round-off error arises only on agents with non-null valuations). Summarizing the results, we have

Proposition 11 *In the combinatorial allocation problem with homogeneous valuations, for any $\delta > 0$, (i) realizing an approximation ratio higher than $1 - \delta/2$ requires communicating at least $\log \binom{\delta^{-1}+L-1}{L-1}$ bits, and (ii) approximation ratio $1 - \delta/\gamma$ is realized by full revelation of valuations rounded off to multiples of δ, which takes at most $N\delta^{-1} \log(L+1)$ bits.*

Corollary 3 *In the combinatorial allocation problem with homogeneous valuations, (i) TPAS in $\log L$ is impossible even for $N = 2$ and $\gamma = 1$, but (ii) full revelation of rounded off valuations is a FPAS in parameters $\log L$, N, and γ^{-1}.*

Proof. (i) If there existed a TPAS in $\log L$, then realizing approximation ratio $1 - 1/(4L)$ would take only polynomial communication in $\log L$. By Proposition 11(i), however, it requires communicating at least $\log \binom{2L+L-1}{L-1} = \log \frac{(3L-1)!}{(2L)!(L-1)!} \geq L - 1$ bits.

(ii) By Proposition 11(ii), full revelation of valuations rounded off to multiples of $\varepsilon\gamma$ realizes approximation ratio $1 - \varepsilon$ using at most $N\varepsilon^{-1}\gamma^{-1} \log(L+1)$ bits. ■

By Proposition 11(ii) means, in particular, that we can achieve any given positive approximation error using $O(\log L)$ bits. As L grows, this amount of communication is proportional to that of simply announcing an allocation (which takes roughly $N \log(L+1)$ bits). On the other hand, Corollary 3(i) means that the extra communication burden needed to guarantee halving the approximation error (regardless of the starting error) is exponential in $\log L$. Intuitively, this means that when the number L of objects is large, and we have a protocol that achieves a close approximation of efficiency, a small reduction in inefficiency requires an enormous increase in communication.

Our analysis can also be related to the model of Calsamiglia (1977), in which instead of L indivisible goods there is one unit of an infinitely divisible good. In this case, V is the space of nondecreasing functions $v : [0,1] \to [0,1]$, and so $\dim V = \infty$. Corollary 1 then implies that efficiency requires an infinite-dimensional message space, re-deriving

Calsamiglia's (1977) result.[19,20] However, Calsamiglia's model allows a FPAS, provided that the agents' valuation functions satisfy

$$|v(x') - v(x)| \leq (-\log|x' - x|)^{-A} \text{ for all } x, x' \in [0,1]$$

for some $A > 0$. Under this mild strengthening of continuity (for example, implied by Hölder continuity of any degree), restricting agents to consume the good in $L = 2^{(\varepsilon\gamma)^{-1/A}}$ identical discrete units reduces the surplus by at most $N(\log L)^{-A} = N\varepsilon\gamma$. Running the protocol described in Corollary 3(ii) on this discretized allocation space will approximate the maximum surplus within $2N\varepsilon\gamma$, and so realize approximation ratio $1 - 2\varepsilon$, while communicating only $N\varepsilon^{-1}\log(L+1) \approx N\varepsilon^{-1}(\varepsilon\gamma)^{-1/A}$ bits. Thus, we have a FPAS, even though exact efficiency in this case requires infinite-dimensional communication.

7.3 Substitute Valuations

Here each agent's valuation space $V_i = V_{\text{sub}}$ is the set of valuations $v \in V_{\text{gen}}$ whose indirect utility function $w(p) = \max_{S \subset L}\left(v(S) - \sum_{l \in S} p_l\right)$ is submodular in $p \in \mathbb{R}^L_+$. This is one of the many equivalent definitions of the substitute property—see Gul and Stacchetti (1999) and Milgrom (2000).[21]

Since $V_{\text{sub}} \subset V_{\text{sm}}$ (see Gul and Stacchetti (1999)), the dual (4) of a substitute valuation is not one, except when both are *additive*, i.e., take the form $v(S) = \sum_{l \in S} \phi_l$ for some $\phi \in \mathbb{R}^L$. Let V_{add} denote the class of additive valuations. Since $V_{\text{add}} \subset V_{\text{sub}}$, and the dual of an additive valuation is itself, Corollary 1 yields

[19] Calsamiglia (1977) restricts the valuation of agent 1 to be concave and that of agent 2 to be convex. Since the dual of a concave valuation is convex, the analysis goes through without modification. Similarly, the agents' valuations can be restricted to be arbitrarily smooth, since smoothness is preserved under duality.

[20] In contrast, when both agents' valuations are known to be concave, a Walrasian equilibrium with a single real-valued price exists and realizes efficiency (regardless of whether the good is divisible or not).

[21] The property is more widely known as "gross substitutes," which is redundant because in the asbence of wealth effects the concepts of gross and net substitutability coincide.

Proposition 12 *In the combinatorial allocation problem with additive or substitute valuations, the dimension of the message space in any efficient protocol is at least* $\dim V_{add} = L$.

This lower bound is attained by the Walrasian equilibrium with per-object prices, which always exists with substitute valuations (Kelso and Crawford 1982, Gul and Stacchetti 1999).

A major disadvantage of the Walrasian protocol is that it is nondeterministic, leaving open the question of how to *find* an equilibrium. Deterministic protocols achieving this were proposed by Gul and Stachetti (2000) and Ausubel (2002). These protocols are variations on the ascending-bid auction with prices quoted for individual items, and so they are only FPAS. This is in fact true of all proposed approximation protocols based on the primal-dual schema (see, e.g., Bikhchandani et al. (2001)).

We improve upon the proposed auction designs by describing a TPAS for this setting. For this purpose, we write the efficient allocation problem as an integer programming problem, letting $x_{iS} = 1$ if agent i's allocation $k^{-1}(i) = S$ and $x_{iS} = 0$ otherwise. As shown by Bikhchandani and Mamer (1997), if a Walrasian equilibrium exists then any efficient allocation must also solve the relaxed surplus-maximization program in which fractional allocations x_{iS} are allowed:

$$\max_{x \in \mathbb{R}_+^{N \cdot 2^L}} \sum_{i \in N, S \subset L} x_S^i v_i(S) \tag{P}$$

$$\begin{aligned} \text{s.t.} \sum_{i \in N, S \ni l} x_{iS} &\leq 1 \text{ for all objects } l \in L, \\ \sum_{S \subset L} x_{iS} &\leq 1 \text{ for all agents } i \in N. \end{aligned}$$

Linear program (P) has only $N+L$ constraints but an exponential number of variables, so it would be hard to solve it directly. It is easier to solve the dual program

$$\min_{p \in \mathbb{R}_+^L, w \in \mathbb{R}_+^N} \sum_{l \in L} p_j + \sum_{i \in N} w_i \tag{D}$$

$$\text{s.t. } w_i - \left[v_i(S) - \sum_{S \ni l} p_l \right] \geq 0 \text{ for all } i \in N,\ S \subset L,$$

where p_l and w_i denote the Lagrange multipliers with the constraints in (P) associated with object l and agent i respectively. Examination of the complementary slackness condition shows that a vector $(x, p, w) \in \mathbb{R}^{N \cdot 2^L} \times \mathbb{R}^L \times \mathbb{R}^N$ is comprised of solutions to (P) and (D) if and only if it describes a Walrasian equilibrium, with x being the (possibly fractional) equilibrium allocation, p the price vector, and w the vector of agents' utilities. We proceed under the assumption that an integral Walrasian equilibrium allocation exists.

While (D) has an exponential number of constraints, each of the constraints depends on the valuation of a single agent. This allows to solve (D) with a separation-based linear programming algorithm, such as the ellipsoid method (see, e.g., Karloff (1991)). The method uses an oracle, who, presented with a candidate solution, produces a violated inequality whenever one exists. Consider a protocol running a separation-based algorithm, but instead of each oracle query, asking each agent i to report a bundle S that gives him a higher net utility than his "utility target" w_i at the current price vector p. If such a report is made by one of the agents, the protocol continues. It is known that when the inputs (valuations) are discrete multiples of δ, the separation-based algorithm produces a solution within a number of steps that is polynomial in the number of variables (in our case $N + L$) and $\log \delta^{-1}$ (the size of each "input" number). Since at each step there are at most N numbers and bundles announced by the agents, the whole protocol uses polynomial communication in N, L, and $\log \delta^{-1}$. Thus, we have a polynomial procedure to calculate the value of δ-discretized program (P), which approximates the true value of (P) to within $N\delta/2$. An approximate integer solution to (P) can then be deduced using standard computational techniques of self-reduction, yielding a TPAS.[22]

[22] Indeed, since (P) has an integer solution, there exists an allocation of item 1 that does not reduce its value, and so does not reduce the value of the δ-discretized (P) by more than $N\delta$. Thus, let us find an agent such that upon allocating item 1 to him, the value of the δ-discretized (P) does not fall by more

8 Average-case Analysis

Suppose that we are given a probability distribution over the states of the world. Then we can relax the notion of approximation to the requirement that only the expected surplus be close to optimal. At the same time, we can count the expected rather than worst-case number of bits transmitted, which allows a savings from coding more frequent messages with fewer bits (as in Shannon's (1948) information theory). The communication measure that uses such average-case counting is called "distributional communication complexity," since the results clearly depend on the assumed joint distribution of the states of the world. For example, if the distribution puts all weight on a single state, then an efficient outcome is known and can be implemented with no communication. Thus, it is only interesting to consider distributions that are sufficiently diffuse so that no outcome has a high a priori probability of being efficient. It turns out that for *some* such distributions, the communication complexity of approximating efficient combinatorial allocations still grows exponentially with the number of objects:

Proposition 13 *In the combinatorial allocation problem, there exists a sequence of joint probability distributions over valuation profiles* $\left(v^1, \ldots, v^N\right)$ *for each* N *and* L *such that for any* $\varepsilon > 0$*, realizing fraction* $1/N + \varepsilon$ *of the maximum expected surplus requires transmitting an expected number of bits that is at least* $c \exp\left\{L/\left(2N^2\right) - 5 \ln N\right\}$*, for some fixed* $c > 0$*.*

Proof. Consider the set packing problem described in the proof of Proposition 7. In the Appendix we prove a lower bound on the communication complexity of distinguishing between the states in which N subsets can be packed from those in which only one subset can be packed (and each agent's individual packing number is at most 1), which applies to randomized protocols with any bounded error. Using the equivalence of randomized

than $N\delta$. Such allocation of item 1 may not be exactly optimal, but it will not reduce the value of (P) by more than $2N\delta$. Then allocate item 2 in the same fashion, then item 3, etc. Since we try allocating each item to each agent, we will use NL calls to the polynomial procedure solving the δ-discretized (P). Since the accumulated loss of surplus is at most $L \cdot 2N\delta$, we have a TPAS.

complexity and distributional complexity (which follows from the Minimax Theorem—see Kushilevitz and Nisan (1997, Section 3.4)), it follows that for some sequence of distributions over states, our lower bound applies to the "Discrimination Problem" of distinguishing the states with surplus $S(v) = N$ from those with $S(v) = 1$ correctly with probability at least $1/2 + \varepsilon/4$. In particular, we must have $\Pr\{S(v) = N\} \leq 1/2 + \varepsilon/4$, for otherwise declaring "N" would solve the Discrimination Problem.

Now consider the conditional distribution on states with $S(v) = N$ (assigning probability zero to all other states), so that the maximum expected surplus is N. Any protocol Γ that achieves fraction $1/N + \varepsilon$ of it on the conditional distribution must realize a surplus greater than 1 with probability at least ε. We can adapt Γ to solve the Discrimination Problem for the original distribution as follows: Run Γ and ask the agents to announce their utilities at the realized allocation. Declare "N" if the sum of the announcements exceeds 1, declare "1" otherwise. The probability of error is at most $\Pr\{S(v) = N\}(1 - \varepsilon) \leq (1/2 + \varepsilon/4)(1 - \varepsilon) < 1/2 - \varepsilon/4$, hence the protocol solves the Discrimination Problem. This implies that Γ must satisfy our lower bound, which is stated in the Proposition. ■

Since the bundled auction guarantees share $1/N$ of the expected surplus, the Proposition implies that for some joint distribution over the agents' valuations, achieving a higher expected surplus than the bundled auction still requires expected communication that is exponential in L.

In the distribution constructed in the above proposition, the valuations are not necessarily independently distributed. We can obtain a (weaker) lower bound on approximation for independently distributed valuations using the distributional lower bounds of Babai, Frankl, and Simon (1986):

Proposition 14 *In the combinatorial allocation problem with $N = 2$ agents, there exists a sequence of probability distributions pairs D_1, D_2 over valuations for each L such that realizing fraction c of the maximum expected surplus (for some fixed $c < 1$) when the agents' valuations are distributed independently according to D_1, D_2, respectively, requires*

communication of an expected number of bits that is exponential in L.

Proof. We will use a reduction to the "disjointness problem" from communication complexity theory (See Kushilevitz and Nisan 1997). In this problem, each agent $i = 1, 2$ is given a subset X_i a set M, and the objective is to decide whether $X_1 \cap X_2 = \emptyset$. Babai, Frankl, and Simon (1986) prove a lower bound on the distributional complexity of disjointness for product distributions:

Theorem 1 (Babai, Frankl, and Simon (1986)) *There exists a distribution D on subsets of M with $|M| = m$, and a fixed $d > 0$, such if the two agents' sets are drawn according to D, then any protocol that communicates in expectation at most $d\sqrt{m}$ bits must err with at least 1% probability when attempting to solve the disjointness problem.*

We will now show that any protocol for combinatorial allocation that achieves 99.5% expected efficiency when the agents' valuations are drawn independently according to distributions D_1, D_2 (to be defined below) can be used to obtain a protocol for the disjointness problem for $m = \binom{L}{L/2}$ that errs on at most 1% of inputs drawn according to D. Thus the lower bound of $d\sqrt{m} = d\sqrt{\binom{L}{L/2}}$ (which is exponential in L) communication applies to the combinatorial allocation problem.

Distributions D_1, D_2 of the two agents' valuations are defined as follows: Let $M = \{S \subset L : |S| = L/2\}$, hence $|M| = m$. For each $i = 1, 2$, the valuation v_i of agent i is chosen by first choosing a random subset $X_i \subset M$ according to the distribution D in the above theorem. We define $v_i(S) = 0$ for $|S| < L/2$; $v_i(S) = 1$ for $|S| > L/2$. For agent 1, we define for $|S| = L/2$, $v_1(S) = 1$ if $S \in X_1$ and $v_1(S) = 0$ otherwise. For agent 2, we define for $|S| = L/2$, $v_2(S) = 1$ if $L \backslash S \in X_2$ and $v_2(S) = 0$ otherwise. Now in order to solve the disjointness problem on X_1 and X_2, the two parties can each create a valuation according to the rule specified above, solve the combinatorial allocation problem, and declare that X_1 and X_2 are not disjoint when the obtained allocation has value 2, and disjoint otherwise. Observe that finding an allocation with surplus 2 means finding a partition of L into two sets $(S, L \backslash S)$ of size $L/2$ each such that $S \in X_1$ and $S \in X_2$,

thus proving that X_1 and X_2 are not disjoint. Thus, the declaration of non-disjointness is always correct. On the other hand, since any inefficient allocation loses at least half the available surplus, an allocation protocol that loses at most 0.5% of expected surplus cannot produce an inefficient allocation with probability more than 1%. Therefore, the probability of falsely declaring disjointness is at most 1%. ■

The proof is done for $c = 99.5\%$, which is derived from a constant quoted in Babai, Frankl, and Simon (1986). No optimization of the constant was attempted and it seems likely that a substantial strengthening is possible.

9 Comparison with Computational Complexity

The communication problem examined here is different from the previously considered problem of *computing* an efficient or approximately efficient allocation when all the valuations are known. The computational complexity of a problem is defined relative to its input size, but in the combinatorial allocation problem the size of the input—a description of the valuations—is itself exponential in the number L of items. For this reason, the computational complexity literature has tended to focus on cases in which the input size is small, such as that of "single-minded preferences," in which each agent values only a single bundle of items. Even in such simple cases from the viewpoint of communication, the efficient combinatorial allocation problem has been shown to be NP-complete.

Nevertheless, we believe that the communication bottleneck is more severe in practice than the computational one. Recall that NP-completeness only indicates that the problem may be exponential asymptotically as the number L of items goes to infinity, and that only if $P \neq NP$, which is considered likely but not proven. In practice, computational complexity can be handled for up to hundreds of items (and thousands of bids) optimally (Vohra and de Vries 2002, Sandholm et al. 2001) and thousands of items (with tens of thousand of bids) near-optimally (Zurel and Nisan 2001). In contrast, we derive *exact* lower bounds on communication complexity for any given L. For example, with general valuations, Proposition 5 establishes that exact efficiency requires communicat-

ing at least one price for each of the $2^L - 1$ possible bundles of objects, and Corollary 6 establishes that with two bidders any improvement over the bundled auction still requires communicating at least $\binom{L}{L/2}$ bits. For example, with $L = 50$ items, any improvement would require the bidders to send at least $\binom{50}{25} \simeq 1.3 \times 10^{14}$ bits $\simeq$ 500 Gigabytes of data, which roughly corresponds to 250 million typewritten pages.[23]

A model of computational complexity that is designed to handle large inputs is "black box complexity." In the general model, the algorithm can ask a "black box" (oracle) an arbitrary query about an agent's valuation v^i. (More restricted models define which queries the black box for v^i will answer—e.g., in the "valuation oracle" model only allows the requests to report of the agents' valuations v^i_k for particular outcomes k.) It is easy to see that the number of queries in the general black box model is bounded below by the deterministic communication complexity of the problem.

10 Incentives

So far we have ignored the agents' incentives to follow the prescribed strategies. If the agents behave in their self-interest, the designer faces additional "incentive-compatibility" constraints requiring that the agents' strategies constitute an equilibrium of the communication game. In this section, we show how in the quasilinear case, these constraints may be satisfied using monetary transfers.

Suppose that after running the protocol, we ask each agent to report his payoff u_i at the resulting allocation, and pay each agent i a transfer $t_i = \sum_{j \neq i} u_i$. This transfer scheme (first proposed by Reichelstein (1984, pp.45-46)) ensures that each agent's total payoff equals the total surplus, and so converts the communication game into one of common interest. (In the terminology of Marschak and Radner (1972), the agents become a "team"). If the protocol is efficient, then obeying the prescribed strategies constitutes

[23] Note also that while the computational burden may be distributed among the bidders, e.g., by asking them to suggest allocations or matches to their package (Banks et al. 1989, Nisan and Ronen 2000), all the communication in a combinatorial auction passes through the auctioneer.

an *ex post equilibrium* under the described transfer scheme: no unilateral deviation by an agent can increase his payoff in any state.[24]

But what if the protocol is not exactly but only approximately efficient? The behavior of rational agents in such a protocol will depend on their beliefs. Let us make the standard assumption in economics that the agents have a common prior over the states of the world, and that they play a Bayesian-Nash Equilibrium (BNE) of the game defined by the protocol. Since the game is one of common interest under the proposed transfer scheme, the strategy profile maximizing the expected surplus constitutes a BNE of the game. In particular, this BNE cannot have a lower expected surplus than that achieved by the original protocol. In other words, if the agents coordinate on this BNE, it will achieve the best average-case approximation consistent with the game.[25] Of course, this relies heavily on the agents' rationality—both individual (being able to calculate an optimal strategy profile) and collective (being able to coordinate on it). But if agents are not fully rational, it is not clear how to consider their incentives in the first place.

A possible criticism of the proposed transfer scheme is that it is very costly. The cost can be covered with lump-sum participation fees, but these fees may be restricted by the agents' participation constraints. The largest fee that still guarantees agent i's participation is the surplus that could be achieved in his absence. (This makes an agent indifferent about participating when he does not contribute anything to the surplus, and strictly prefer to participate otherwise.) Together with the transfer scheme, this gives each agent his Vickrey-Groves-Clarke (VCG) payoff. As noted by Reichelstein (1984) and Ausubel (2002), calculation of the participation fees requires running the protocol with successively removing each agent, which multiplies the communication burden by

[24] In general, obedience will not be a dominant strategy, since an agent i may gain from a deviation if he expects another agent j to use a strategy that is not consistent with any type u_j.

[25] To take this observation to its extreme, under our transfer scheme, rational agents need not be offered a protocol at all! Namely, given an explicit expression for communication costs, the agents could be made to internalize these costs (e.g., by paying for sending bits). Then if the agents play some "free form" game in which they can individually send messages and implement an allocation, the protocol that maximizes the expected surplus net of communication costs will constitute a BNE of the game.

$N+1$. However, some efficient protocols (such as those in Feigenbaum et al. (2000) and Bikhchandani et al. (2001)) yield the VCG transfers as a side product, which provides a substantial communication savings when N is large.

11 Conclusion

Price mechanisms are the most commonly observed and the best studied economic allocation mechanisms. However, until now there has not been a complete understanding of their role. To be sure, the Welfare Theorems show that the Walrasian price mechanism produces efficient allocations in convex economies. Still, the possibility remained that other mechanisms also produce efficient allocations in convex economies, or that non-price mechanisms perform better than price mechanisms in nonconvex economies. The designers of combinatorial auctions have proposed numerous designs that purport to find efficient allocations without finding all the prices supporting them.

The present paper has shown that in fact, *any* efficient mechanism is "essentially" a price mechanism, in the sense that it must reveal supporting prices along with the efficient allocation itself. Thus, the indispensable role of prices for implementing efficient allocations is now made clear. This result holds regardless of the agents' incentives, even if the agents report truthfully.

In the combinatorial allocation problem, efficient communication must name (at least) one price for each of the $2^L - 1$ possible bundles, where L is the number of objects. We demonstrate that all possible monotonic price vectors must be used, hence the required communication is at least as extensive as full revelation of one agent's preferences. Even if we only require a better approximation of efficiency than that obtained by auctioning off all objects as a bundle, we must still use at least $\binom{L}{L/2}$ bits in a two-bidder auction. With $L = 50$ (a realistic number), this amount of information roughly corresponds to more than 250 million typewritten pages. We also show that even if approximation is required only on expectation, for *some* probability distribution over valuations, it still requires exponential communication in L. These results imply that for realistic values of

L, any combinatorial auction design or other "preference elicitation" scheme suggested in the literature would either run for a prohibitively long time, or, if stopped after some reasonable time, would not produce an efficient or even approximately efficient allocation.

Our results should not be taken to imply that all real-life combinatorial auctions are useless, any more than Arrow's impossibility theorem implies that all real-life institution are useless. Rather, by showing that no institution is guaranteed to achieve good results on the universal preference domain, Arrow's theorem has led researchers to examine the performance of specific institutions on restricted domains. Similarly, by showing that there does not exist a practical auction design that works well for all possible combinatorial preferences over many objects, we hope to motivate auction designers to focus on specific classes of preferences or probability distributions over them. However, the burden should be on the proposer of a particular design to characterize the environments on which it works well. The tools developed in the present paper will be useful for this purpose, as we have demonstrated by examining the communication burden for the cases of submodular, homogeneous, and substitute valuations.

Finally, we have clarified the validity of measuring the communication burden of efficiency with the dimension of the required message space, as is common in the economic literature. The key question is whether this measure accurately reflects the difficulty of approximating efficiency with a discretized mechanism. We find that the dimensionality of message space is indicative of the complexity of achieving a "truly polynomial" approximation of efficiency. On the other hand, a somewhat slower but still practical "fully polynomial" approximation is sometimes achieved with much less communication. In such cases, the economic measure may seriously overstate the "hardness" of the communication problem. A dramatic example of this is offered by Calsamiglia's (1977) model of allocating a homogeneous divisible good, in which exact efficiency requires infinite-dimensional communication, but we demonstrate a fully polynomial approximation mechanism.

Appendix: A Set Packing Lower Bound

Our lower bound for the set packing problem will use a reduction from the following "approximate-disjointness" problem which was studied in Alon et al. (1999):

The Approximate Disjointness Problem: Each player $i \in N$ holds a subset $B_i \subset T$ of a finite set T. We are required to distinguish between the following two extreme cases:

- Negative: $\cap_{i \in N} B_i \neq \emptyset$,
- Positive: for every $i \neq j$, $B_i \cap B_j = \emptyset$.

A lower bound on the required communication of cT/N^4 for some fixed constant $c > 0$ was given in Alon et al. (1999) for randomized protocols (with two-sided error). The lower bound was improved by Jaikumar Radhakrishnan and Venkatesh Srinivasan to $\ln 2 \cdot T/N$ for deterministic and nondeterministic protocols.

Let us now define the approximate set packing problem. Each agent $i \in N$ agents holds a collection $A_i \subset 2^L$. The objective is to approximate the *packing number*—the number of subsets in $\cup_{i \in N} A_i$ that are packed together, i.e., are pairwise disjoint. We will prove a lower bound on the communication complexity of distinguishing between the case where the packing number is 1 (i.e, any two sets $S_i \in A_i$ and $S_j \in A_j$ for $i, j \in N$ intersect) and the case where there exist N disjoint sets $S_1 \in A_1, ..., S_N \in A_N$, and so the packing number is N.

We reduce this problem from the approximate-disjointness problem on a family T of partitions of set L into N subsets. That is, each $t \in T$ is a partition $(t_1, \ldots, t_N)$ of L. We require family T to have the following property:

Definition 4 *A family T of partitions of set L among N agents has the pairwise-intersection property if for every $t', t'' \in T$ such that $t' \neq t''$ we have $t'_i \cap t''_j \neq \emptyset$ for every $i, j \in N$, i.e., any two elements of different partitions intersect.*

Lemma 2 *There exists a family of partitions of set L among N agents with the pairwise-intersection property of size $e^{L/(2N^2)}/N$.*

Proof. We use the probabilistic method: Construct each partition $t = 1, \ldots T$ randomly, by randomly and independently placing each element of L with equal probability in one of the parts of t. Perform this construction independently for all $\dot{t}$. Now, for given $i, j \in N$ and $t', t'' = 1, ..., T$ such that $t' \neq t''$,

$$\Pr\left\{t'_i \cap t''_j = \emptyset\right\} = \prod_{l \in L} \Pr\left\{l \notin t'_i \cap t''_j\right\} = \left(1 - 1/N^2\right)^L \leq e^{-L/N^2}.$$

The probability that $t'_i \cap t''_j = \emptyset$ for *some* $i, j \in N$ and $t', t'' = 1, ..., T$ such that $t' \neq t''$ is at most N^2T^2 times that number. Therefore, when $N^2T^2e^{-L/N^2} < 1$, with a positive probability we obtain a family of partitions with the pairwise-intersection property, hence such a family must exist. ■

We can now specify the reduction of approximate disjointness on a set T to the approximate set packing problem. Set T is taken to represent a family of partitions of L among N with the pairwise intersection property. Agent i receiving as input a set $B_i \subset T$ constructs the collection $A_i = \{t_i | t \in B_i\}$. Now, if there exists $t \in \cap_i B_i$, then $(t_1, \ldots, t_N)$ constitutes an N-packing. If, on the other hand, $B_i \cap B_j = \emptyset$ for all $i \neq j$, then for any two sets $t'_i \in A_i$ and $t''_j \in A_j$ we have $t' \neq t''$ and thus by the pairwise intersection property, $t'_i \cap t''_j \neq \emptyset$ for every $i, j \in N$, hence the packing number is at most 1.

From the lower bounds described above for the approximate disjointness problem we obtain the following lower bounds for the set packing problem:

Theorem 2 *Any N-player protocol (deterministic or nondeterministic) realizing an approximation ratio higher than $1/N$ for the set packing problem communicates at least $\ln 2 \cdot e^{L/(2N^2) - 2\ln N}$ bits. A randomized protocol achieving this communicates, on expectation, at least $ce^{L/(2N^2) - 5\ln N}$ bits, for some fixed $c > 0$.*

References

[1] Alon, N., Y. Matias, and M. Szegedy (1999). "The Space Complexity of Approximating the Frequency Moments," *Journal of Computer and System Sciences*, 58(1); 137-137.

[2] Ausubel, L. (2000). "An efficient dynamic auction for heterogeneous commodities," mimeo, University of Maryland.

[3] Ausubel, L., and P. Milgrom (2002). "Ascending Auctions with Package Bidding," *Frontiers of Theoretical Economics* 1(1).

[4] Babai, L., P. Frankl, and J. Simon. "Complexity Classes in Communication Complexity Theory," *Proceedings of FOCS*, 337–347, 1986.

[5] Banks, J., J. Ledyard, and D. Porter (1989). "Allocating Uncertain and Unresponsive Resources: An Experimental Approach," *Rand Journal of Economics*, 20:1 – 25.

[6] Bikhchandani, S., and J.W. Mamer (1997). "Competitive equilibrium in an exchange economy with indivisibilities," *Journal of Economic Theory*, 74:385 – 413.

[7] Bikhchandani, S., R. Vohra, S. de Vries, and J. Schummer (2001). "Linear programming and Vickrey auctions," mimeo, 2001. Availailabe from http://www.kellogg.nwu.edu/faculty/vohra/htm/res.htm.

[8] Calsamiglia, X. (1977). "Decentralized Resource Allocation and Increasing Returns," *Journal of Economic Theory* 14: 262-283.

[9] Edgar, G.E. (1990). *Measure, Topology, and Fractal Geometry*. New York: Springer-Verlag.

[10] Feigenbaum, J. , C. H. Papadimitriou, and S. Shenker (2000). "Sharing the cost of multicast transactions," STOC 2000.

[11] Gul, F., and E. Stacchetti (1999). "Walrasian equilibrium with gross substitutes," *Journal of Economic Theory*, 87:95 – 124.

[12] Gul, F., and E. Stacchetti (2000). "The English Auction with Differentiated Commodities," *Journal of Economic Theory*.

[13] Hayek, F. (1945). "The Use of Knowledge in Society," *American Economic Review* 35: 519-30.

[14] Holzman, R., N. Kfir-Dahav, D. Monderer, and M. Tennenholtz (2001), "On Bundling Equilibrium in Combinatorial Auctions" mimeo, Technion University, http://iew3.technion.ac.il/~moshet/rndm11.ps

[15] Hurwicz, L. (1960), "Optimality and Informational Efficiency in Resource Allocation Processes," in K.J. Arrow, S. Karlin, and P. Suppes, eds., *Mathematical Methods in the Social Sciences*, 27-46, Stanford: Stanford University Press.

[16] Hurwicz, L. (1977). "On the Dimensional Requirements of Informationally Decentralized Pareto-Satisfactory Processes," in K.J. Arrow and L. Hurwicz, eds., *Studies in Resource Allocation Processes*, 413-424, New York: Cambridge University Press.

[17] Hurwicz, L., and T. Marschak (2003a), "Finite Allocation Mechanisms: Approximate Walrasian versus Approximate Direct Revelation," *Economic Theory* 21, 545-572.

[18] Hurwicz, L., and T. Marschak (2003b), "Comparing Finite Mechanisms," *Economic Theory* 21, 783-841.

[19] Jordan, J.S. (1982), "The Competitive Allocation Process is Informationally Efficient Uniquely," *Journal of Economic Theory* 28(1), 1-18.

[20] Jordan, J.S., and D. Xu (1999), "On the Communication Complexity of Expected-Profit Maximization," *Journal of Economic Theory* 86, 185-202.

[21] Karloff, H. (1991). *Linear Programming*. Birkhäuser Verlag.

[22] Kelso, A.S. Jr., and V.P. Crawford (1982). "Job Matching, Coalition Formation, and Gross Substitutes." *Econometrica*, 50, 1483-1504.

[23] Korshunov, A.D. (1981). "O Chisle Monotonnykh Bulevykh Funktsi*i", Problemy Kibernetiki 38, 5-108.

[24] Kushilevitz, E., and N. Nisan (1997). *Communication Complexity*. Cambridge University Press.

[25] Kwasnica, A., J. Ledyard, D. Porter, and C. DeMartini (2002). "A new and improved design for multi-object iterative auctions," available from http://www.ices-gmu.net/pdf/materials/372.pdf

[26] Lang, S. (1987). *Linear Algebra*, New-York: Springer-Verlag, 3rd edition.

[27] Lehmann, D., L. Ita O'Callaghan, and Y. Shoham (1999). "Truth revelation in rapid, approximately efficient combinatorial auctions." *1st ACM conference on electronic commerce*.

[28] Lehmann, B., D. Lehmann, and N. Nisan (2001). "Combinatorial Auctions with Decreasing Marginal Utilities." In *ACM conference on electronic commerce*.

[29] Marschak, J., and R. Radner (1972), *Economic Theory of Teams*, New Haven: Yale University Press.

[30] Marschak, T. (1996) "On Economies of Scope in Communication," *Economic Design* 2(1): 1-30.

[31] Milgrom, P. (2000). "Putting Auction Theory to Work: The Simultaneous Ascending Auction," *Journal of Political Economy*.

[32] Mount, K., and S. Reiter (1974). "The Information Size of Message Spaces," *Journal of Economic Theory* 28:1-18.

[33] Mount, K., and S. Reiter (1977). "Economic Environments for which there are Pareto Satisfactory Mechanisms," *Econometrica* 45:821-842.

[34] Nisan, N. (2000). "Bidding and Allocation in Combinatorial Auctions." In *ACM Conference on Electronic Commerce*.

[35] Nisan, N., and A. Ronen (2000). "Computationally Feasible VCG-Based Mechanisms." In *ACM Conference on Electronic Commerce*.

[36] Parkes, D.C. (1999). "ibundle: An Efficient Ascending Price Bundle Auction," mimeo, Harvard University.

[37] Parkes, D.C. (2002). "Price-Based Information Certificates for Minimal-Revelation Combinatorial Auctions," in *Agent-Mediated Electronic Commerce IV*, Padget et al. (eds), LNAI 2531 pp. 103-122, Springer-Verlag.

[38] Parkes, D.C., and Lyle H. Ungar (2000). Iterative combinatorial auctions: Theory and practice. In *AAAI/IAAI*, pages 74–81.

[39] Reichelstein, S. (1984). "Incentive Compatibility and Informational Requirements," *Journal of Economic Theory* 34(1), 32-51.

[40] Rothkhof, M.H., A. Pekeč, and R.M. Harstad (1998). "Computationally manageable combinatorial auctions." *Management Science*, 44(8):1131–1147.

[41] Sandholm, T., S. Suri, A. Gilpin, and D. Levine "Cabob: A fast optimal algorithm for combinatorial auctions." In *IJCAI*, 2001.

[42] Sato, F. (1981). "On the Informational Size of Message Spaces for Resource Allocation Processes in Economies With Public Goods," *Journal of Economic Theory*, 24, 48-69

[43] Shannon, C.E. (1948). "A Mathematical Theory of Communication," *Bell System Technical Journal* 27, 379-423, 623-656.

[44] Tian, G. (2001). "A Unique Informationally Efficient Allocation Mechanism in Economies with Public Goods," mimeo, Texas A&M University.

[45] Tian, G. (2004). "A Unique Informationally Efficient Allocation Mechanism in Economies with Consumption Externalities," *International Economic Review* 45 (1), pp. 79-111.

[46] Van Zandt, T. (1999), "Decentralized Information Processing in the Theory of Organizations," in Murat Sertel, ed., *Contemporary Economic Issues, Vol. 4: Economic Design and Behavior*. London: MacMillan Press Ltd., pp. 125-160.

[47] Vazirani, V.V. (2001), *Approximation Algorithms*, Berlin: Springer-Verlag.

[48] Vohra, R., and S. de Vries (2002). "Combinatorial Auctions: A Survey," *INFORMS Journal of Computing*, forthcoming.

[49] Walker, M. (1977). "On the Informational Size of Message Spaces," *Journal of Economic Theory* 15, 366-375.

[50] Yao, A. C.-C. (1979), "Some complexity questions related to distributive computing," In *ACM Symposium on Theory of Computing*, pp. 209–213.

[51] Zinkevich, M., Blum, A., T. and Sandholm (2003), "Preference Elicitation and Allocation with Read-Once Preferences.," to appear in *2003 ACM conference on Electronic Commerce*.

[52] Zurel, E., and N. Nisan (2001). "An efficient approximate allocation algorithm for combinatorial auctions," *ACM conference on 2001 Electronic Commerce*.

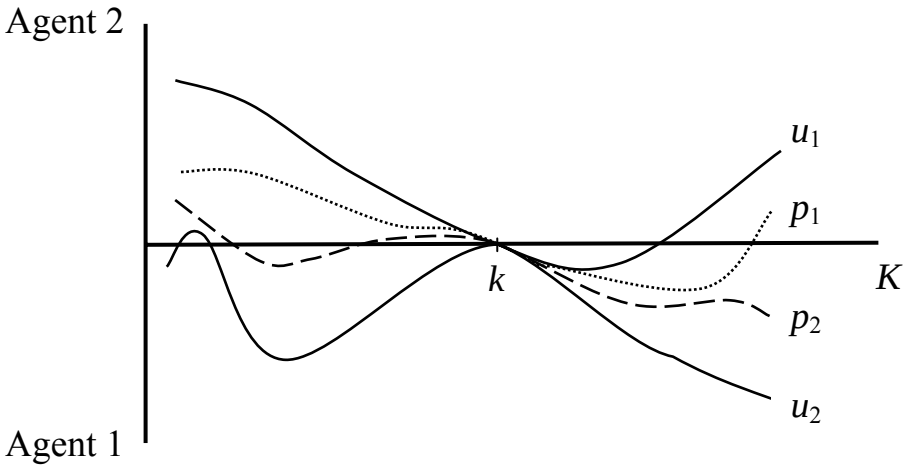

Figure 1: Price Equilibrium

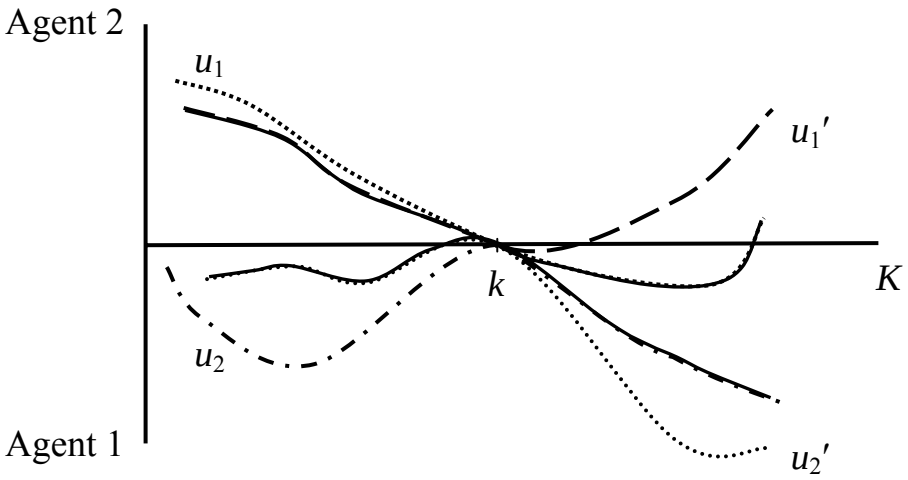

Figure 2: Proof of Proposition 1