The Efficiency-Incentive Tradeoff in Double Auction Environments

Gabriel Carroll, Department of Economics, MIT
gdc@mit.edu

December 10, 2011

Abstract

We consider the tradeoff between efficiency and incentives in large double auction environments with weak budget balance. No mechanism simultaneously gives agents perfect incentives to be truthful and ensures first-best efficiency, but a planner designing a mechanism may be willing to compromise on either of these dimensions for improvements along the other. She would then naturally wish to find where the possibility frontier lies with respect to incentives and efficiency. We make inroads on this question: our main result locates the frontier to within a factor that is logarithmic in the size of the market.

Thanks to (in random order) Alessandro Bonatti, Xiao Yu Wang, Ruitian Lang, Glenn Ellison, Parag Pathak, Rakesh Vohra, and Juuso Toikka for helpful comments and advice.

1 Introduction

1.1 Overview

Economists have known since Akerlof [2] that private information can prevent markets from reaching efficient outcomes. Moreover, the results of Myerson and Satterthwaite [25], among many others, show that this inefficiency is not specific to competitive markets but rather is unavoidable under any possible mechanism for allocating goods. However, some mechanisms lead to more severe inefficiency than others, and so the natural next question is what second-best mechanism achieves outcomes that are as efficient as possible. A large literature addresses this question in many different settings.

Customarily, the mechanism design literature assumes that agents optimize perfectly. In particular, applying the revelation principle, it is standard to take as a given constraint that each agent's best possible strategy should be to truthfully reveal his private information, and then describe the optimal mechanism subject to this constraint.

However, in practice, human decision-makers are not perfectly strategic, or at least do not perfectly optimize the material payoffs that are usually modeled. Accordingly, a planner could offer a mechanism asking agents to report their preferences, in which reporting truthfully is not exactly optimal, but the incentives to behave strategically instead are small. The planner might then expect that agents will report truthfully, rather than go to the trouble of figuring out how to strategically manipulate the mechanism. This notion leads to a tradeoff between incentives and efficiency, and motivates a quantitative examination of the tradeoff.

The present paper makes initial inroads into quantitatively studying this tradeoff, in the specific context of large double auction environments with quasilinear preferences and weak budget balance. This is one of the most widely studied economic environments for mechanism design, and can be viewed as an analytically convenient, stylized model of an exchange economy.

By studying the incentive-efficiency tradeoff, we bridge two branches of theoretical research on mechanisms for large markets. On one hand is the literature, going back to Roberts and Postlewaite [27], showing that in large exchange economies, under the competitive equilibrium mechanism, the incentives for strategic misreporting of preferences (assuming other agents are truthful) go to zero. On the other hand is a recent literature studying exact equilibria of large markets and showing that the inefficiency goes to zero [16, 17, 28]. In particular, part of that latter literature [13, 29] takes a mechanism design approach and identifies the optimal rate at which *any* mechanism can converge to full efficiency as the market becomes large. However, no previous work has explored the space in between these branches, looking for compromises between perfect efficiency and perfect incentives. If it turned out that large gains in efficiency could be achieved at the cost of a very small relaxation of incentives, that would cast a new light on the existing convergence-rate bounds. Conversely, if this were not possible, the existing impossibility results would be strengthened.

Our modeling framework is fundamentally non-equilibrium-based, intended to study design of market institutions for agents who are not perfectly familiar with their environment. Indeed, our basic motivating assumption — that agents do not effortlessly know how to manipulate to their advantage — would be difficult to justify in an equilibrium model. On the other hand this assumption is reasonable for describing plenty of exchange in real-world markets. The typical shopper at the grocery store is unlikely to think about the demand curve of other shoppers for a pint of strawberries, or to know how he might profitably deviate from pure price-taking behavior so as to influence the prices he faces — or to even want to bother thinking about how he might go about strategically deviating.

To explore quantitatively the tradeoff between incentives and efficiency, we need ways to measure both. Following [8], we work in a direct revelation framework, where a mechanism asks each agent his value for the good being exchanged, and determines trades accordingly; and we take a worst-case approach to the definition of incentives. The *susceptibility to manipulation* of a particular market mechanism is the largest amount of expected utility any agent could possibly gain by reporting his value strategically instead of truthfully; the maximum is taken over all possible beliefs about the distributions from which other agents' behavior is drawn. Likewise, we also use a worst-case measure for inefficiency: it is the largest value, over all possible distributions of agents' valuations, of the expected shortfall in surplus realized by the mechanism compared to the first-best (assuming that agents report truthfully).[1]

Our worst-case methodology is appropriate for a planner choosing a trading institution to be used in the future, when she does not have clear priors over agents' valuations or their strategic behavior, and wants to be sure that her mechanism will perform well. (The author's other work [8] fleshes out in detail a positive model of such a planner's choice of mechanism, showing how our measurement methodology fits in.) In addition, when defining susceptibility, note that we take the worst case over *beliefs*: there is no presumption that agents know the true distribution of others' behavior. This is in keeping with our non-equilibrium framework, in which agents may not accurately know the details of their environment.

Two mechanisms in the existing literature represent polar cases with respect to the efficiency-incentive tradeoff. On one end is the *k-double auction*, a version of the competitive mechanism; where the goods are given to the traders whose (reported) values are highest, and trades take place at a market-clearing price. This mechanism achieves

[1] We measure inefficiency using the allocation of goods, not the sum of the agents' utilities. These measures are different if the mechanism runs a surplus. Our measure implicitly assumes that the surplus can be paid to someone outside the mechanism.

first-best efficiency if traders are truthful, but does not provide perfect incentives for truthfulness. On the other end is McAfee's [23] dominant-strategy double auction, which provides perfect incentives, but may fail to realize (at most) one profitable trade.

Our results, presented in Section 3, describe the asymptotic behavior of susceptibility or inefficiency as the number of agents becomes large. We consider environments in which buyers' valuations are independently drawn from one distribution, sellers' valuations are independently drawn from another distribution, and these two distributions are not too dissimilar. More precisely, the distributions have densities that differ everywhere by at most some fixed ratio. Then the k-double auction has susceptibility on the order of $1/\sqrt{N}$, and McAfee's double auction has inefficiency on the order of $1/\sqrt{N}$, where N represents the size of the market. Our main result (Theorem 3) shows that both mechanisms are close to the possibility frontier: There is a constant c such that any mechanism has either susceptibility or inefficiency at least $c/(\sqrt{N}\log N)$.

The assumption of similar distributions is necessary. If we allow the buyers' and sellers' valuations to come from arbitrarily different distributions, then the susceptibility-or-inefficiency lower bound does not go to zero as the market grows (Proposition 4).

In Section 4, we address a possible "consequentialist" critique of our methodology: Perhaps a planner designing mechanisms should not be concerned with incentives for strategic manipulation per se, since agents might manipulate in a way that does not adversely affect the outcome of the mechanism. Instead, she should be concerned with the inefficiency that will result from manipulation. It turns out that our results withstand this critique, as long as we make reasonably conservative assumptions about how agents might try to manipulate. Specifically, we allow that agents may attempt any manipulation that gives them a sufficiently large gain in expected utility (they will not necessarily find the optimal manipulation), and we consider the inefficiency that may result. In this formulation, instead of a tradeoff between efficiency under truthfulness and incentives for truthfulness, we have a tradeoff between efficiency under manipulation and the planner's confidence about agents' cost of strategic behavior. Only a little extra work is needed to reformulate our main results in these terms.

In addition to the results themselves, the method of proof for the lower bounds merits attention. We use a straightforward variation on a standard proof of the impossibility of attaining both first-best efficiency and perfect incentives. That proof uses the usual integral formula derived from the envelope theorem to compute the utility that each type of each agent would need to receive, and verifies that the total surplus in the market is not enough to provide that utility to each agent. We introduce error terms into the proof, representing inefficiency and susceptibility to manipulation. By continuity, the same contradiction is still reached if the error terms are sufficiently small; we simply track them explicitly to find out how large they need to be to avoid a contradiction. Some care is needed in working the error terms into the integral formula: it turns into a discrete approximation, and one needs to choose the approximation points appropriately. However, the fact that we can readily adapt a standard argument to obtain our results is encouraging, since it suggests that similar methods can be applied to study tradeoffs involving incentives in other mechanism design domains.

1.2 Literature review

The question of incentives for truthfulness in large markets can be traced back to Roberts and Postlewaite [27], who showed that the benefits from misreporting one's demand function in an exchange economy (under the Walrasian mechanism) go to zero as the economy is replicated. More recent work in the market design literature gives similar results for matching mechanisms [3, 15, 18, 19], argues that this property makes the mechanisms suitable for use in practice. A variety of other literature has also considered mechanisms with small incentives to manipulate [7, 12, 20, 21, 22, 24, 30], but without looking at the possibility frontier between these incentives and other properties of the mechanism, as we emphasize here.

In contrast to this approach, much of the recent work on double auctions has assumed that agents perfectly optimize — thus imposing Bayesian Nash equilibrium, with given valuation distributions — and examined either behavior in specific mechanisms or the design problem of finding the optimal mechanism. Several relevant papers studied rates of convergence to perfect efficiency. In the model of Rustichini, Satterthwaite, and Williams [28], equilibrium behavior in the k-double auction leads to inefficiency tending to zero as the market grows, at rate $1/N$. McAfee's dominant-strategy double auction [23] also attains rate $1/N$. Satterthwaite and Williams [29] showed (for the uniform distribution) that any mechanism has inefficiency of order at least $1/N$, so that the the two mechanisms just described are asymptotically optimal, to within a constant factor. (These results appear to contradict our Theorem 3 below, which implies worse rates of convergence. The discrepancy arises because we allow for a broader class of value distributions.) There is also recent work on large double auctions with interdependent values, e.g. [26]. However, our focus here is on environments with private values.

2 Model

2.1 Elements

We consider double auction settings with unit capacity, private values, and quasilinear utility. Thus, there are N sellers who each have a good to sell, and N buyers who each would like to buy a good. Write b_i for the value of the good to buyer i, and s_i for the value to seller i. These values are normalized to lie in $[0,1]$. We write P^b, P^s for profiles of buyers' and sellers' valuations, $(b_i)_{i=1,\ldots,N}$ and $(s_i)_{i=1,\ldots,N}$, and $P = (P^b, P^s)$ for the profile of all $2N$ agents' valuations. Then P^b_{-i} denotes the profile of valuations of all buyers except the ith, and P^s_{-i} similarly.

We focus attention on direct mechanisms. (This, and other assumptions, will be discussed in Subsection 2.2.) Thus, a *mechanism* elicits each agent's valuation, and determines an allocation of the goods (possibly probabilistic) and expected transfer payments as a function of the reported valuations. Formally, a mechanism is a collection of $4N$ functions,

$$M = (p^b_i, p^s_i, t^b_i, t^s_i)_{i=1,\ldots,N}$$

where

$$p^b_i, p^s_i : [0,1]^{2N} \to [0,1]$$

denote each agent's probability of exchange (i.e. p^b_i is buyer i's probability of receiving a good, and p^s_i is seller i's probability of giving up a good); and

$$t^b_i, t^s_i : [0,1]^{2N} \to \mathbb{R}$$

denote the net payment made by each agent. We require the functions $p^b_i, p^s_i, t^b_i, t^s_i$ to be measurable. We also impose the feasibility conditions

$$\sum_i p^b_i(P) = \sum_i p^s_i(P)$$

for every profile of valuations $P \in [0,1]^{2N}$.

We do not allow the mechanism to run a deficit, but we do allow a surplus; thus we impose weak budget balance:

$$\sum_i t^b_i(P) + t^s_i(P) \geq 0$$

for all P. (With deficits allowed, the Vickrey-Clarke-Groves mechanism [10, 14] would achieve full efficiency in dominant strategies, so the tradeoff between efficiency and incentives would be uninteresting.)

If the profile of reported valuations is P, then the utilities of buyer i and seller i, respectively (relative to not participating in the mechanism), are

$$U_i^b(P) = b_i p_i^b(P) - t_i^b(P), \qquad U_i^s(P) = -s_i p_i^s(P) - t_i^s(P).$$

In addition to feasibility and weak budget balance, we also require mechanisms to satisfy ex post individual rationality:

$$U_i^b(P), U_i^s(P) \geq 0$$

for all profiles P and all i. Note that individual rationality and weak budget balance imply that the transfers $t_i^b(P), t_i^s(P)$ are bounded.

In the operation of a mechanism, we assume that the buyers' valuations are drawn independently from a distribution F^b on $[0, 1]$, and the sellers' valuations are drawn independently from a distribution F^s. We will in general not presume these distributions are known, either to the planner or to the agents, but rather allow a set $\mathcal{F}$ of possible pairs (F^b, F^s). We will assume that for all possible pairs, F^b, F^s are representable by bounded density functions on $[0, 1]$. Our results would be unchanged (and indeed simpler to prove) if we allowed for atoms in the distributions, but by requiring continuity we make clear that atoms are not driving the results. We will sometimes write f^b, f^s for the respective density functions.

The utility achieved by buyer i when the *reported* profile is $\widehat{P}$ but his true valuation is b_i is

$$U_i^b(\widehat{P}|b_i) = b_i p_i^b(\widehat{P}) - t_i^b(\widehat{P}).$$

Similarly define

$$U_i^s(\widehat{P}|s_i) = -s_i p_i^s(\widehat{P}) - t_i^s(\widehat{P}).$$

We define the *susceptibility to manipulation* of a mechanism M as the strongest possible incentive faced by any agent to misreport his valuation. Formally, for a given set $\mathcal{F}$ of distribution pairs, the *buyer-susceptibility* is

$$\sigma^b = \sup_{i, b_i, \widehat{b}_i, (F^b, F^s)} \Big(E_{(F^b, F^s)}[U_i^b(\widehat{b}_i, P_{-i}^b, P^s | b_i)] - E_{(F^b, F^s)}[U_i^b(P|b_i)] \Big)$$

where the supremum is over buyers i, true valuations $b_i \in [0, 1]$, possible reports $\widehat{b}_i \in [0, 1]$, and distribution pairs $(F^b, F^s) \in \mathcal{F}$. The expectations are with respect to other agents' reported types, where we assume other buyers' reports are drawn from F^b and sellers' from F^s (all independently). Similarly the *seller-susceptibility* is

$$\sigma^s = \sup_{i, s_i, \widehat{s}_i, (F^b, F^s)} \left(E_{(F^b, F^s)}[U_i^s(P^b, \widehat{s}_i, P_{-i}^s | s_i)] - E_{(F^b, F^s)}[U_i^s(P | s_i)] \right).$$

The *susceptibility* is then

$$\sigma = \max\{\sigma^b, \sigma^s\}.$$

The motivating story behind this definition is simple: Suppose a planner knows that agents face a psychological or computational cost of at least ϵ to behaving strategically. If the planner chooses a mechanism whose susceptibility is known to be less than ϵ, then agents will not bother to behave strategically and instead will simply report their true valuations. This is discussed in more detail in [8], which also shows how the above definition of susceptibility is equivalent to one in which players are allowed to be uncertain about the distribution pair (F^b, F^s).

We define the *inefficiency* of a mechanism using an analogous worst-case formulation. For any profile P of valuations, define the first-best welfare $W^{FB}(P)$ to be the sum of the N highest valuations, and the welfare $W^M(P)$ achieved by the mechanism as $\sum_i b_i p_i^b(P) + \sum_i s_i(1 - p_i^s(P))$. Note that

$$W^M(P) = \left[\sum_i U_i^b(P) + \sum_i U_i^s(P) \right] + \left[\sum_i t_i^b(P) + \sum_i t_i^s(P) \right] + \left[\sum_i s_i \right].$$

The second bracketed expression is the surplus accrued by the mechanism; we implicitly assume when computing welfare that this surplus can be paid to an outside agent. The third expression is independent of the choice of mechanism, so it does not affect the shortfall relative to first best, $W^{FB}(P) - W^M(P)$.

The inefficiency of M relative to $\mathcal{F}$ is then defined as

$$\sup_{(F^b, F^s)} \left(E_{(F^b, F^s)}[W^{FB}(P) - W^M(P)] \right)$$

where the supremum is over $(F^b, F^s) \in \mathcal{F}$, and the expectation is with respect to valuation profiles where each b_i is drawn from F^b and each s_i is drawn from F^s (independently). In particular, this definition of inefficiency assumes truthful reporting; we will address this issue in Section 4. Also, the definition is absolute (not normalized by the size of the market), though our results could just as well be formulated in terms of relative inefficiency.

We will be mainly concerned with a set of distribution pairs $\mathcal{F}$ in which the buyers' and sellers' value distributions are not too different. Specifically, let $\lambda \geq 1$ be an exogenously given constant; then define $\mathcal{F}_\lambda$ to be the family of distribution pairs (F^b, F^s) whose densities satisfy $f^b(x)/\lambda \leq f^s(x) \leq \lambda f^b(x)$ for all $x \in [0,1]$. (As a special case, $\lambda = 1$ means that the buyers' and sellers' values are drawn from the same distribution.) Our main results apply to $\mathcal{F}_\lambda$. However, we will also consider the set $\mathcal{F}_\infty$, of all possible pairs (F^b, F^s) of distributions representable by bounded density functions on $[0,1]$.

Note that we have not required mechanisms to be *anonymous* — that is, to treat all buyers and all sellers identically. Formally, a mechanism M is anonymous if, for all profiles (b_i, s_i) and all permutations π^b, π^s of $\{1, \ldots, N\}$, we have

$$p_i^b(b_{\pi^b(1)}, \ldots, b_{\pi^b(N)}, s_{\pi^s(1)}, \ldots, s_{\pi^s(N)}) = p^b_{\pi^b(i)}(b_1, \ldots, b_N, s_1, \ldots, s_N)$$

for each i, and similarly for the functions p_i^s, t_i^b, t_i^s. However, to study the inefficiency-susceptibility frontier, it is enough to consider anonymous mechanisms. Indeed, if M is any mechanism with susceptibility σ and inefficiency η, we can define an anonymous mechanism $\widetilde{M}$ by randomly permuting the buyers and the sellers and then applying M: that is,we define

$$\widetilde{p}_i^b(b_1, \ldots, b_N, s_1, \ldots, s_N) = \frac{1}{(N!)^2} \sum_{\pi^b, \pi^s} p^b_{(\pi^b)^{-1}(i)}(b_{\pi^b(1)}, \ldots, b_{\pi^b(N)}, s_{\pi^s(1)}, \ldots, s_{\pi^s(N)})$$

and define $\widetilde{p}_i^s, \widetilde{t}_i^b, \widetilde{t}_i^s$ likewise; these comprise the mechanism $\widetilde{M}$. Then $\widetilde{M}$ is an average of $(N!)^2$ mechanisms, all of which (by symmetry) have gains at most σ to any agent from manipulating and all of which have an expected welfare loss at most η relative to the first-best, so the same is true of $\widetilde{M}$. Thus we have an anonymous mechanism whose susceptibility and inefficiency are at most those of M.

Given this, we will henceforth restrict attention to anonymous mechanisms without further comment.

2.2 Discussion

There are a couple of assumptions implicit in the above modeling framework which call for elaboration. Our restriction to direct mechanisms really entails two assumptions: first, that each agent's strategy depends only on his valuation (and no other information); second, that the strategy space can be taken to be the space of valuations, with honest reporting as the default behavior of agents who do not strategize.

The second assumption is actually not a serious restriction. We view double auction environments as a stylized model of competitive markets, and truthfulness as a metaphor for price-taking. This seems a natural assumption about default behavior (especially for inexperienced participants). But more generally, we could take an indirect-mechanism approach, allowing a mechanism M to specify any strategy space for each player, together with probabilities of trade and transfers as functions function of the strategy profile, *and* a specification of a default strategy for each player (possibly mixed) that depends on that player's valuation. By a straightforward variation of the usual revelation principle, M could be converted into a direct mechanism M', where default behavior consists of honest reporting, and where M' has the same inefficiency as M and susceptibility no higher than M (it may have strictly lower susceptibility, due to the elimination of strategies in M that were not default strategies for any type). Since we are concerned only with the inefficiency-susceptibility frontier, it suffices to focus on direct mechanisms as we have done above.

The assumption that players' behavior depends only on their valuations is more serious. This assumption invites the critique of Bergemann and Morris [6] that a planner could potentially do better by designing a mechanism in which agents also condition their strategies on their beliefs about other agents' behavior. If we were to formulate the mechanism design problem in full generality taking this into account, a direct mechanism would have agents report their full types, where a type consists not only of a valuation but an entire belief hierarchy (including beliefs about any parameters relevant to agents' manipulative behavior — see the discussion in Section 4 below).

However, recall that we have chosen to make no assumptions about the correctness of agents' beliefs about others' behavior. The appropriate worst-case measure of inefficiency in this framework would specify that for a mechanism to have inefficiency at most η, the expected welfare loss relative to first-best should be at most η for every possible distribution of buyer and seller *types*, regardless of whether or not their beliefs reflected the true distribution. With such a definition, it turns out that our results would remain valid even in this more fully-specified setting. This is because the proofs of our lower bounds rely only on a single "true" distribution pair (F^b, F^s) when analyzing the incentive to misreport, and so these lower bounds actually hold for the subset of the type space on which it is common knowledge among the agents that values are drawn from this (F^b, F^s).

On this subset, two types of a given agent differ only in their valuation, so assuming that agents report only their valuations is without loss of generality. (However, the proofs do analyze *inefficiency* using distributions other than the fixed (F^b, F^s), so it is crucial that the definition of inefficiency allows the "true" distribution to be differ from the one that agents believe to be correct.) Thus the critique of [6] does not bind here. The formal details of this argument would be notationally involved and not relevant to the main point of this paper, so we omit them.

We should also comment here on the interpretation of the individual rationality constraints, which we have written in an ex post form. These can be thought of as normative constraints on acceptable mechanisms. They can also be viewed in positive terms, if agents have the opportunity to renege after the mechanism has operated. However, this latter interpretation is less tidy: as pointed out by Compte and Jehiel [11], the proper formulation of such a constraint is as a *veto constraint*, which not only requires ex post individual rationality but also imposes stronger incentive constraints — agents should not be able to benefit by misreporting their valuation and then potentially vetoing the outcome depending on the realizations of other agents' types. This distinction turns out to be immaterial for our results, however: our negative results under individual rationality still hold a fortiori under the stronger veto constraint, and it can be checked that our positive results also hold, since the relevant mechanisms (the McAfee and k-double auctions) satisfy the veto constraint.

Alternatively, using a richer type space as outlined above, in which strategies reflect an agent's full type, would allow us to instead use an interim version of the individual rationality constraints — each agent has nonnegative expected utility from participation — in which case the positive interpretation would be straightforward. Our lower bounds wouls still hold with these weaker constraints rather than the ex post constraints, again for the reason that the proofs invoke the constraints only for agents whose beliefs coincide with the true distributions (F^b, F^s). Again, we omit the details.

2.3 Polar mechanisms

We now describe in precise terms our two polar mechanisms. We will content ourselves with verbal descriptions, rather than tediously write out all the algebraic expressions.

For any $k \in [0, 1]$, the *k-double auction* (described e.g. in [28]) is as follows. For any profile P of $2N$ reported valuations, sort them as $v_{(1)} \geq v_{(2)} \geq \cdots \geq v_{(2N)}$, and define the price $p^* = kv_{(N)} + (1-k)v_{(N+1)}$. Allocate the goods to the agents with the

N highest valuations. (If there is a tie at $v_{(N)}$, ration uniformly at random; ties are not really important since they occur with probability zero in our model.) Every buyer who receives a good pays p^*, and every seller who sells a good receives p^*. It is clear that this mechanism satisfies feasibility, budget balance, and individual rationality, and that it achieves inefficiency of 0.

McAfee's double auction, from [23], is a bit more complex. The rules are as follows. Sort the buyers' reported valuations in decreasing order, $b_{(1)} \geq \cdots \geq b_{(N)}$, and the sellers' in increasing order, $s_{(1)} \leq \cdots \leq s_{(N)}$. Also define $b_{(0)} = s_{(N+1)} = 1$ and $b_{(N+1)} = s_{(0)} = 0$ for convenience. Let k be the highest value satisfying $b_{(k)} \geq s_{(k)}$; this is the efficient number of trades. We have $0 \leq k \leq N$. Define the price $p^* = (b_{(k+1)} + s_{(k+1)})/2$.

If $p^* \in [s_{(k)}, b_{(k)}]$, then have the k highest-value buyers buy the good from the k lowest-value sellers at price p^*. (Again, break ties uniformly at random.) Otherwise, note that $k > 0$, and have the $k-1$ highest-value buyers each receive a good and pay $b_{(k)}$, while the $k-1$ lowest-value sellers each sell their good for price $s_{(k)}$. The mechanism thus carries out $k-1$ trades and earns a budget surplus $(k-1)(b_{(k)} - s_{(k)}) \geq 0$.

This mechanism is again feasible, weakly budget-balanced, and individually rational. It has been established that reporting truthfully is a dominant strategy for all agents in this mechanism [23, Theorem 1]. Therefore, it has a susceptibility of 0.

3 The efficiency-incentive tradeoff

We can now properly introduce our results on the efficiency-incentive tradeoff.

The results are illustrated in Figure 1, where the gray region represents the (inefficiency, susceptibility) pairs (η, σ) attained by some mechanism. The frontier must be convex, as shown in the figure: If mechanism M has inefficiency η and susceptibility σ, and mechanism M' has inefficiency η' and susceptibility σ', then for any $\alpha \in [0, 1]$ we can take the convex combination $(1-\alpha)M + \alpha M'$ (defined by taking corresponding convex combinations of the $p_i^b, p_i^s, t_i^b, t_i^s$ functions), and this mechanism has inefficiency at most $(1-\alpha)\eta + \alpha\eta'$ and susceptibility at most $(1-\alpha)\sigma + \alpha\sigma'$.

For the main results, we consider the class of distribution pairs $\mathcal{F}_\lambda$, in which some similarity is imposed between the buyers' and sellers' value distributions. We give the approximate locations of the two polar mechanisms, which lie on the two axes of the possibility set, at a distance of order $1/\sqrt{N}$ from the origin. On the other hand, we identify a point lying below the possibility set (indicated by the star in the figure), whose coordinates are of order $1/(\sqrt{N} \log N)$. Thus, these results together pin down the location

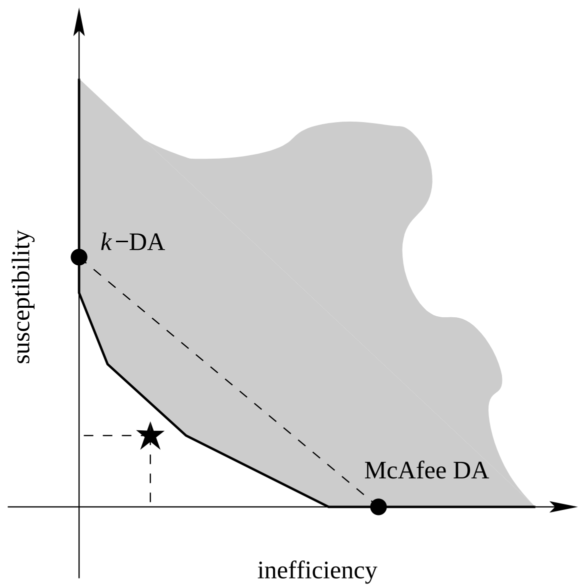

Figure 1: The possibility frontier

of the possibility frontier to within a factor that is logarithmic in the size of the market.

If we look at the class of distribution pairs $\mathcal{F}_\infty$, where the distribution of buyers' values can be arbitrarily different from the distribution of sellers' values, then a similar picture applies but on a different scale: the lower bound on inefficiency or susceptibility (the star point) does not go to zero as the market becomes large. This will be shown in Subsection 3.2.

3.1 Main results

We first bound the inefficiency attained by the McAfee double auction, over $\mathcal{F}_\lambda$. As the number of agents grows, the inefficiency shrinks on the order of $N^{-1/2}$. More specifically:

Proposition 1 *There is a constant c such that the McAfee double auction has inefficiency at most $c/\sqrt{N}$ on $\mathcal{F}_\lambda$. (The value of c depends on λ.)*

The calculation is routine, but rather lengthy, so we leave it for Appendix A. For a quick overview: Inefficiency is at most the value of the least valuable trade; a change-of-variables argument implies that this value is no greater than the probability that the least valuable trade involves of a buyer with value above x^* and a seller with value below x^*, for a suitable (fixed) x^*. For this to happen, in turn, it must be that either (a) the number of agents with values above x^* is close to N, which happens with probability on the order of

$N^{-1/2}$ by a law-of-large-numbers argument; or (b) when all $2N$ agents are arranged from highest value to lowest, there is a long run of consecutive buyers or consecutive sellers, which happens with probability decreasing exponentially in the length of the run.

We can also bound the susceptibility of the k-double auction; it is also on the order of $N^{-1/2}$. This is because the probability that any given misreport is pivotal — that is, that it advantageously changes the market price — is of order at most $N^{-1/2}$, by a central-limit-theorem argument.

Proposition 2 *There is a positive constant c such that the k-double auction has susceptibility at most $cN^{-1/2}$. (Again, c may depend on λ.)*

Proof: Consider a buyer with value b, reporting a false value $\widehat{b}$. We may assume $\widehat{b} < b$, since reporting $\widehat{b} > b$ can never be profitable: holding fixed the realizations of other agents' reports, such a misreport cannot decrease the price, nor can it change the buyer's outcome from not receiving a good to receiving one, unless the trade occurs at a price higher than b.

Moreover, again holding fixed the other agents' reports, the misreport can only be beneficial if is pivotal — more specifically, if exactly $N-1$ other agents report values higher than $\widehat{b}$. Indeed, if more than $N-1$ other agents report higher values, then the misreporting buyer gets no good and hence utility zero; if fewer than $N-1$ other agents report higher values, then the misreport has no effect on the price at which he trades.

Since the buyer's realized utility is always between 0 and 1, his expected gain from misreporting is at most the probability that exactly $N-1$ other agents report a value greater than $\widehat{b}$. Letting J be the number of other buyers whose values are less than $\widehat{b}$, we can express this probability as a sum over possible values of J:

$$\sum_{J=0}^{N-1} \binom{N-1}{J} \binom{N}{J} F^b(\widehat{b})^J F^s(\widehat{b})^{N-J} (1-F^b(\widehat{b}))^{N-1-J} (1-F^s(\widehat{b}))^J. \tag{1}$$

We finish by invoking Lemma 9 in Appendix A. If $F^b(\widehat{b}) \leq 1/2$, then using $\binom{N-1}{J} \leq \binom{N}{J}$, the expression in (1) is

$$\leq 2 \sum_{J=0}^{N} \binom{N}{J}^2 F^b(\widehat{b})^J F^s(\widehat{b})^{N-J} (1-F^b(\widehat{b}))^{N-J} (1-F^s(\widehat{b}))^J$$

which, according to the lemma (with $\kappa = 1/2$, say, and $K = 0$), is at most $c\sqrt{\lambda/N}$ for some *absolute* constant c. This certainly implies the desired bound on the buyer's probability of being pivotal.

If $F^b(\widehat{b}) > 1/2$, then using $\binom{N-1}{J} \leq \binom{N}{J+1}$, the expression in (1) is

$$\leq 2 \sum_{J=0}^{N-1} \binom{N}{J+1} \binom{N}{J} F^b(\widehat{b})^{J+1} F^s(\widehat{b})^{N-J} (1 - F^b(\widehat{b}))^{N-1-J} (1 - F^s(\widehat{b}))^J$$

and by a change of variable, this is

$$= 2 \sum_{J=1}^{N} \binom{N}{J-1} \binom{N}{J} (1 - F^s(\widehat{b}))^{N-J} (1 - F^b(\widehat{b}))^{J-1} F^s(\widehat{b})^J F^b(\widehat{b})^{N-J+1}$$

which, again according to the lemma (with $K = 1$), is at most $c\sqrt{\lambda/N}$ for an absolute constant c.

This shows in both cases that the buyer-susceptibility of the k-double auction satisfies the bound. The argument for seller-susceptibility is identical.

□

Having established these estimates for the two polar mechanisms, we can proceed to our main result: a lower bound showing that the two polar mechanisms are close to the optimal rate of convergence of inefficiency or susceptibility as the number of agents becomes large.

Theorem 3 *There exists a positive constant c such that, on $\mathcal{F}_1$, every mechanism has either inefficiency at least $c/(\sqrt{N} \log N)$ or susceptibility at least $c/(\sqrt{N} \log N)$.*

Of course, the same bounds a fortiori hold for any $\mathcal{F}_\lambda$, $\lambda \geq 1$.

The idea behind the proof is as follows: Consider the incentives facing a given agent — say, a buyer — when he believes the other agents' values are drawn from a distribution with mass concentrated near 0 and 1. Let $\overline{p}^b(b)$ be the probability that the buyer gets a good when his value is b (and he reports truthfully). Let $\overline{U}^b(b)$ be the expected utility he attains if his value is b. Similarly define $\overline{p}^s(s)$ and $\overline{U}^s(s)$.

Suppose the mechanism were to have inefficiency and susceptibility zero. Then the first-best allocation would determine $\overline{p}^b$ and $\overline{p}^s$ completely. In turn, these determine the functions $\overline{U}^b$ and $\overline{U}^s$ via the familiar integral formula coming from the envelope theorem (up to a constant, which is bounded below by individual rationality). These expected utility functions are not consistent with weak budget balance — there is not enough expected surplus in the market to give all agent types the needed utility levels.

Let $\overline{p}^{b*}, \overline{p}^{s*}, \overline{U}^{b*}, \overline{U}^{s*}$ be the functions obtained in the above calculations assuming zero inefficiency and susceptibility. With a small amount of wiggle room, we know only that $\overline{p}^b$ and $\overline{p}^s$ have to be close to $\overline{p}^{b*}$ and $\overline{p}^{s*}$, and in turn that $\overline{U}^b, \overline{U}^s$ have to be close to $\overline{U}^{b*}, \overline{U}^{s*}$. Requiring the agents' expected utility levels to be far enough from $\overline{U}^{b*}, \overline{U}^{s*}$ to avoid exceeding the total surplus in the market then leads to a lower bound on either inefficiency or susceptibility.

Proof of Theorem 3: It is enough to prove the result for N sufficiently large; we can then adjust the constant c to ensure the result holds for small N as well.[2]

Suppose that the number c is such that some mechanism M has susceptibility σ and inefficiency η both less than $c/(\sqrt{N}\log N)$. Our goal is to show that c must be larger than some absolute constant. Specifically, we will show that $c \geq 1/7000$. (This is far from best possible, but we are not concerned here with fine-tuning constants.) Thus, suppose that $c < 1/7000$, and seek a contradiction.

Let γ be a sufficiently small positive number. At several points in the course of the proof, we will use the fact that γ is smaller than various functions of N, η, and c. Rather than writing out explicit bounds here, we will simply assume without further comment that all needed bounds are satisfied (there will be only finitely many of them, so this assumption is safe).

Define the density function f by

$$f(x) = \begin{cases} 1/2\gamma, & 0 \leq x \leq \gamma \\ 0, & \gamma < x < 1-\gamma \\ 1/2\gamma, & 1-\gamma \leq x \leq 1. \end{cases}$$

Let F be the corresponding cumulative distribution function. We focus on the incentives facing a given agent when all other agents' reports are independently drawn from F.

Step 1 (buyers). As in the sketch above, let $\overline{p}^b(b)$ be the probability that buyer i receives a good, when his value is b. (By anonymity, this is independent of i.) In this first step, we use efficiency to show that $\overline{p}^b$ is close to its first-best value.

However, because we are working with continuous distributions, efficiency in expectation imposes no restrictions on $\overline{p}^b(b)$ itself for any single value of b. Instead, we need to

[2]To be precise, this requires knowing that for each small N, either inefficiency or susceptibility must be bounded away from 0. By continuity arguments, it is enough to show that there is no mechanism with inefficiency and susceptibility both 0. This can be proven e.g. by using revenue equivalence to show that any such mechanism would have to be equivalent to a VCG mechanism, which always runs a deficit; see [32].

talk about averages. Accordingly, for $\frac{\gamma}{2} \le b \le 1 - \frac{\gamma}{2}$, define

$$\overline{p}^b_\gamma(b) = \frac{1}{\gamma} \int_{b-\frac{\gamma}{2}}^{b+\frac{\gamma}{2}} \overline{p}^b(b')\, db'.$$

We will show that $\overline{p}^b_\gamma(b)$ is approximately bounded below times 1/2 minus a constant times η/b. Specifically, for any $b > 3\gamma/2$,

$$\overline{p}^b_\gamma(b) \ge \frac{1}{2} - \frac{16\eta}{b - 3\gamma/2}. \tag{2}$$

To show this, suppose otherwise, so that

$$\frac{1}{16}\left(b - \frac{3\gamma}{2}\right)\left(\frac{1}{2} - \overline{p}^b_\gamma(b)\right) > \eta. \tag{3}$$

Define a density function g by

$$g(x) = \begin{cases} \frac{1}{\gamma}, & b - \frac{\gamma}{2} < x < b + \frac{\gamma}{2} \\ 0 & \text{otherwise.} \end{cases}$$

Define the density $h(x) = \left(1 - \frac{1}{N}\right) f(x) + \left(\frac{1}{N}\right) g(x)$. Let G, H be the distributions associated with g, h.

Suppose that we draw all $2N$ agents' values independently from H. This is equivalent to generating values as follows: we mark each agent as an F-type or G-type agent, randomly with probability $1 - \frac{1}{N}$ or $\frac{1}{N}$ respectively, and then draw the valuations from F or G accordingly. Let E denote the event that there is exactly one G-type buyer and no G-type seller. We have

$$Pr(E) = \left[N \cdot \left(1 - \frac{1}{N}\right)^{N-1} \cdot \left(\frac{1}{N}\right)\right] \cdot \left(1 - \frac{1}{N}\right)^{N} \ge \frac{1}{16}. \tag{4}$$

Conditional on E, the G-type buyer receives a good with probability $\overline{p}^b_\gamma(b)$.

On the other hand, conditional on E, all F-type agents have values distributed uniformly on the set $[0, \gamma] \cup [1 - \gamma, 1]$. In this case, the probability that at least half the F-type agents have values in $[0, \gamma]$ is 1/2, by symmetry. Thus, conditional on E, we have probability at least 1/2 that the G-type buyer is among the top N values, and the next lower value is at most γ. In particular, conditional on E, there is probability at least $1/2 - \overline{p}^b_\gamma(b)$ that the G-type buyer is among the top N values but does not receive a good, and the next highest value is at most γ. When this occurs, there is an efficiency loss (relative to first-best) of at least $b - \frac{3\gamma}{2}$.

Therefore, conditional on E, we have an expected efficiency loss (relative to first-best) of at least $\left(\frac{1}{2} - \overline{p}^b_\gamma(b)\right)\left(b - \frac{3\gamma}{2}\right)$. Since $Pr(E) \geq 1/16$, we have an unconditional inefficiency of at least $\frac{1}{16}\left(\frac{1}{2} - \overline{p}^b_\gamma(b)\right)\left(b - \frac{3\gamma}{2}\right)$. But this amount is greater than η by (3). We have a contradiction. Therefore, (2) must hold.

In fact, we have the simpler bound

$$\overline{p}^b_\gamma(b) \geq \frac{1}{2} - \frac{32\eta}{b} \tag{5}$$

(as long as $\gamma < 64\eta/3$). Indeed, (2) implies (5) for $b \geq 3\gamma$, and for $b < 3\gamma$ the right side of (5) is negative, so the inequality holds trivially.

Henceforth we will only need this latter bound.

Step 2 (buyers). We next construct a discrete approximation for the standard integral formula, leading to a lower bound on the utilities of buyer types with high values. Specifically, we will show that buyers with values in the interval $[1-\gamma, 1]$ must, on average, achieve utility at least $1/2 - 1/20\sqrt{N}$.

To this end, let $\overline{t}^b(b)$ be the expected payment by buyer i, when his value is b, and other values are drawn independently from F. Again, this is independent of i. Let $\overline{U}^b(b) = b\overline{p}^b(b) - \overline{t}^b(b)$ be the expected utility achieved by a buyer with value b.

Take $K = \lfloor \log N \rfloor$. Define buyer values $b_0, b_1, \ldots, b_K$ by

$$b_j = \left(1 - \frac{\gamma}{2}\right)^{1-\frac{j}{K}} \left(\frac{1}{20\sqrt{N}}\right)^{\frac{j}{K}}.$$

(The subscripts here simply index the values; they do not denote different buyer identities.) These buyer values will essentially serve as the interval endpoints in our approximation to the integral formula. However, instead of using these values exactly, we will need to average over small perturbations of the values. This is because our available bounds on probabilities of trade apply to the averages $\overline{p}^b_\gamma$, not to $\overline{p}^b$ for any single type.

Define ρ to be the ratio of successive b_j's:

$$\rho = \frac{b_j}{b_{j+1}} = \left(\frac{1 - \frac{\gamma}{2}}{1/20\sqrt{N}}\right)^{1/K}$$

and note that

$$\rho \leq (20\sqrt{N})^{1/K} \leq 20^{2/\log N}(\sqrt{N})^{2/\log N} = 20^{2/\log N} e < 3 \tag{6}$$

(as long as N is large enough, as usual).

Now, by definition of σ, for any $r \in [-\gamma/2, \gamma/2]$, a buyer of type $b_j + r$ (for any j) cannot benefit by more than σ from misreporting as type $b_{j+1} + r$.

Consider any such r . We have

$$\begin{aligned}
\overline{U}^b(b_j + r) &= (b_j + r)\overline{p}^b(b_j + r) - \overline{t}^b(b_j + r) \\
&\geq (b_j + r)\overline{p}^b(b_{j+1} + r) - \overline{t}^b(b_{j+1} + r) - \sigma \\
&= \overline{U}^b(b_{j+1} + r) + (b_j - b_{j+1})\overline{p}^b(b_{j+1} + r) - \sigma
\end{aligned}$$

for each j. By combining these inequalities for each j we get

$$\begin{aligned}
\overline{U}^b(b_0 + r) &\geq \overline{U}^b(b_K + r) + \sum_{j=0}^{K-1}(b_j - b_{j+1})\overline{p}^b(b_{j+1} + r) - K\sigma \\
&\geq \sum_{j=0}^{K-1}(b_j - b_{j+1})\overline{p}^b(b_{j+1} + r) - K\sigma \qquad (7)
\end{aligned}$$

where the last step is by individual rationality.

Now average over $[-\gamma/2, \gamma/2]$. For each j, we know from (5) that

$$\frac{1}{\gamma}\int_{-\frac{\gamma}{2}}^{\frac{\gamma}{2}} \overline{p}^b(b_{j+1} + r)\, dr \quad \geq \quad \frac{1}{2} - \frac{32\eta}{b_{j+1}}.$$

Therefore,

$$\begin{aligned}
\frac{1}{\gamma}\int_{-\frac{\gamma}{2}}^{\frac{\gamma}{2}} \overline{U}^b(b_0+r)\,dr &\geq \sum_{j=0}^{K-1}(b_j-b_{j+1})\left[\frac{1}{\gamma}\int_{-\frac{\gamma}{2}}^{\frac{\gamma}{2}}\overline{p}^b(b_{j+1}+r)\,dr\right]-K\sigma \\
&\geq \sum_{j=0}^{K-1}(b_j-b_{j+1})\left(\frac{1}{2}-\frac{32\eta}{b_{j+1}}\right)-\frac{c}{\sqrt{N}} \\
&= (b_0-b_K)\left(\frac{1}{2}\right)-K(\rho-1)\,(32\eta)-\frac{c}{\sqrt{N}} \\
&\geq \left(1-\frac{\gamma}{2}-\frac{1}{20\sqrt{N}}\right)\left(\frac{1}{2}\right)-(\log N)\left(64\frac{c}{\sqrt{N}\log N}\right)-\frac{c}{\sqrt{N}} \\
&> \frac{1}{2}-\frac{1}{40\sqrt{N}}-\frac{70c}{\sqrt{N}} \\
&> \frac{1}{2}-\frac{1}{20\sqrt{N}}. \qquad (8)
\end{aligned}$$

(The fourth line uses (6), and the sixth uses the assumption $c < 1/7000 < 1/2800$.)

Now, to wrap up this step of the proof, consider the expected utility accruing to buyer i, when all agents' values are drawn independently from F (and all agents report truthfully). With probability $1/2$, buyer i has a value in the interval $[1-\gamma, 1]$; and conditional on being in this interval, buyer i's value is uniformly distributed on the interval. Therefore, buyer i's unconditional expected utility is at least

$$\frac{1}{2}\left(\frac{1}{2}-\frac{1}{20\sqrt{N}}\right)=\frac{1}{4}-\frac{1}{40\sqrt{N}}.$$

Steps 1, 2 (sellers). The analysis up to this point has focused on incentives for buyers. However, exactly the same calculations can be performed with incentives for sellers. We briefly outline the arguments. Let $\overline{p}^s(s)$ be the probability that a seller with value s sells his good, when all other agents' values are independently drawn from F. Define

$$\overline{p}^s_\gamma(s)=\frac{1}{\gamma}\int_{s-\frac{\gamma}{2}}^{s+\frac{\gamma}{2}}\overline{p}^s(s')\,ds'.$$

We can use the same efficiency arguments as before to obtain a counterpart to (5):

$$\overline{p}^s_\gamma(s)\geq\frac{1}{2}-\frac{32\eta}{1-s}. \qquad (9)$$

Now define $\bar{t}^s(s)$ be the expected net transfer paid by a seller with value s, and let $\overline{U}^s(s) = -s\overline{p}^s(s) - \bar{t}^s(s)$ be the expected utility such a seller attains. Define K as before, and define the seller values $s_0, \ldots, s_K$ by

$$s_j = 1 - \left(1 - \frac{\gamma}{2}\right)^{1-\frac{j}{K}} \left(\frac{1}{20\sqrt{N}}\right)^{\frac{j}{K}}.$$

As before, for any $r \in [-\gamma/2, \gamma/2]$,

$$\overline{U}^s(s_j + r) \geq \overline{U}^s(s_{j+1} + r) + (s_{j+1} - s_j)\overline{p}^s(s_{j+1} + r) - \sigma$$

for each j. Summing over j, averaging over $r \in \left[-\frac{\gamma}{2}, \frac{\gamma}{2}\right]$, and applying (9), we obtain

$$\frac{1}{\gamma}\int_{-\frac{\gamma}{2}}^{\frac{\gamma}{2}} \overline{U}^s(s_0 + r)\, dr > \frac{1}{2} - \frac{1}{40\sqrt{N}} - \frac{70c}{\sqrt{N}} > \frac{1}{2} - \frac{1}{20\sqrt{N}}.$$

Finally, as with the buyers, we conclude that when all agents' values are independently drawn from F, each seller's expected utility is at least $1/4 - 1/40\sqrt{N}$.

Step 3. To complete the proof, we use the lower bound on each agent's utility from Step 2, compare to the total expected surplus available, and obtain a contradiction.

The lower bound of $1/4 - 1/40\sqrt{N}$ obtained at the end of Step 2 holds for each of the $2N$ agents, and therefore the expected surplus generated by the mechanism — that is, the expected sum of the agents' utilities — is bounded below as

$$E[\sum_i U_i^b(P) + \sum_i U_i^s(P)] \geq \frac{N}{2} - \frac{\sqrt{N}}{20}. \tag{10}$$

On the other hand, due to weak budget balance, the surplus at any profile P satisfies

$$\sum_i U_i^b(P) + \sum_i U_i^s(P) \leq \sum_i b_i p_i^b(P) - \sum_i s_i p_i^s(P) \leq W^{FB}(P) - \sum_i s_i. \tag{11}$$

Let's bound the expectation of the first-best welfare W^{FB}. Each agent's value is either in $[0, \gamma]$ or in $[1-\gamma, 1]$, independently with probability $1/2$. Letting K be the number of agents with high values, we can bound the first-best by summing over possible values of

K:

$$\begin{aligned}
E[W^{FB}(P)] &\le \sum_{K=0}^{2N} \binom{2N}{K} \left(\frac{1}{2}\right)^{2N} [\min\{N,K\} \cdot 1 + (N - \min\{N,K\}) \cdot \gamma] \\
&\le N\gamma + \sum_{K=0}^{2N} \binom{2N}{K} \left(\frac{1}{2}\right)^{2N} \min\{N,K\}.
\end{aligned}$$

Break the sum into terms with $K \le N - \lfloor \sqrt{N}/4 \rfloor$ and $K > N - \lfloor \sqrt{N}/4 \rfloor$, rearrange, and then use Lemma 10 from Appendix A (a crude central-limit-theorem approximation) to bound from below the probability that $K \le N - \lfloor \sqrt{N}/4 \rfloor$:

$$\begin{aligned}
E[W^{FB}(P)] &\le N\gamma + \sum_{K=0}^{N-\lfloor\sqrt{N}/4\rfloor} \binom{2N}{K} \left(\frac{1}{2}\right)^{2N} \cdot \left(N - \left\lfloor \frac{\sqrt{N}}{4} \right\rfloor\right) \\
&\qquad + \sum_{K=N-\lfloor\sqrt{N}/4\rfloor+1}^{2N} \binom{2N}{K} \left(\frac{1}{2}\right)^{2N} \cdot N \\
&= N\gamma + \sum_{K=0}^{2N} \binom{2N}{K} \left(\frac{1}{2}\right)^{2N} \cdot N \\
&\qquad - \sum_{K=0}^{N-\lfloor\sqrt{N}/4\rfloor} \binom{2N}{K} \left(\frac{1}{2}\right)^{2N} \cdot \left\lfloor \frac{\sqrt{N}}{4} \right\rfloor \\
&\le N\gamma + N - \frac{1}{4} \left\lfloor \frac{\sqrt{N}}{4} \right\rfloor \\
&< N - \frac{\sqrt{N}}{20}.
\end{aligned}$$

This bounds the expectation of $W^{FB}(P)$.

The expression (11) also involves a $\sum_i s_i$ term. But since each seller has expected value N, the expectation of this sum is simply $N/2$. Consequently, (11) implies that the expected surplus is less than

$$\frac{N}{2} - \frac{\sqrt{N}}{20}.$$

Comparing with (10), we have a contradiction, which completes the proof.

□

3.2 Unrestricted distributions

We now show how the results change when no restrictions are imposed on the pair of distributions — we use the full class $\mathcal{F}_\infty$, rather than $\mathcal{F}_\lambda$

Trivially, the McAfee double auction has inefficiency at most 1 (since it omits at most one desirable trade), and the k-double auction has susceptibility at most 1 (since no agent can ever achieve utility greater than 1). Thus, it is possible to achieve zero inefficiency or susceptibility and a constant, independent of the market size, along the other dimension. The following result shows that it is not possible to do better:

Proposition 4 *There exists a positive constant c such that, on $\mathcal{F}_\infty$, every mechanism has either inefficiency or susceptibility at least c.*

The argument is somewhat similar to that of Theorem 3, but simpler. Since the proof is relatively brief, we will not bother explicitly breaking it into steps.

Proof: We will give a proof with $c = 1/128$. So suppose for contradiction that some mechanism M has susceptibility σ and inefficiency η both less than $1/128$. Let γ be a positive number, chosen to be very small; as in the proof of Theorem 3, we will not bother being explicit about the bounds needed on γ.

Let the distributions F^b, F^s be given by the densities

$$f^b(x) = \begin{cases} 1/\gamma, & 1-\gamma \leq x \leq 1 \\ 0, & 0 \leq x < 1-\gamma \end{cases}$$

$$f^s(x) = \begin{cases} 1/\gamma, & 0 \leq x \leq \gamma \\ 0, & \gamma < x \leq 1. \end{cases}$$

Also let G^b be the distribution with density

$$g^b(x) = \begin{cases} 1/\gamma, & \frac{1}{4}-\gamma \leq x \leq \frac{1}{4} \\ 0 & \text{otherwise,} \end{cases}$$

and take $H^b(x) = \left(1 - \frac{1}{N}\right) F^b(x) + \left(\frac{1}{N}\right) G^b(x)$. Drawing a buyer's value from H^b is equivalent to designating the buyer as F^b-type or G^b-type, with probabilities $1 - \frac{1}{N}$ or $\frac{1}{N}$ respectively, and then drawing the value from F^b or G^b accordingly.

Suppose all buyers' values are drawn from H^b and all sellers' values are drawn from F^s. Let E be the event that there is exactly one G^b-type buyer. By calculations similar to (4), we have

$$Pr(E) \geq \frac{1}{4}.$$

Whenever E occurs, the first-best allocation assigns all the goods to the buyers, and any failure to assign a good to some buyer entails an efficiency loss of at least $(1/4-\gamma)-(\gamma) \geq 1/8$. In particular, if π is the probability that the G^b-type buyer ends up with a good (conditional on E), we have

$$\eta \geq Pr(E) \cdot (1-\pi) \cdot \frac{1}{8} \geq \frac{1-\pi}{32}$$

from which

$$\pi \geq 1 - 32\eta > \frac{3}{4}.$$

Now let $\overline{p}^b(b)$ denote the probability that a buyer receives a good, when he reports b and all other agents' reports are drawn from (F^b, F^s). The above implies that the average of $\overline{p}^b(b)$ with respect to G^b is at least $3/4$:

$$\frac{1}{\gamma}\int_{\frac{1}{4}-\gamma}^{\frac{1}{4}} \overline{p}^b(b)\, db \geq \frac{3}{4}.$$

Let $\overline{U}^b(b)$ be the utility attained by a buyer with value b, when other agents' reports are drawn from (F^b, F^s).

For any $r \in [0, \gamma]$, a buyer of value $1-r$ cannot benefit by more than σ by misreporting as value $1/4 - r$. And so, as in Step 2 of the proof of Theorem 3, we have

$$\overline{U}^b(1-r) \geq \overline{U}^b\left(\frac{1}{4} - r\right) + \left(1 - \frac{1}{4}\right)\overline{p}^b\left(\frac{1}{4} - r\right) - \sigma.$$

Averaging over $r \in [0, \gamma]$ gives

$$\begin{aligned}
\frac{1}{\gamma}\int_{1-\gamma}^{1} \overline{U}^b(b)\, db &\geq \frac{1}{\gamma}\int_{\frac{1}{4}-\gamma}^{\frac{1}{4}} \overline{U}^b(b)\, db + \frac{3}{4}\left[\frac{1}{\gamma}\int_{\frac{1}{4}-\gamma}^{\frac{1}{4}} \overline{p}^b(b)\, db\right] - \sigma \\
&\geq \frac{3}{4}\left[\frac{1}{\gamma}\int_{\frac{1}{4}-\gamma}^{\frac{1}{4}} \overline{p}^b(b)\, db\right] - \sigma \\
&\geq \frac{9}{16} - \sigma \\
&> \frac{1}{2}.
\end{aligned}$$

Now finally suppose all agents' values are drawn from (F^b, F^s). Each buyer's expected utility from the mechanism is $(1/\gamma)\int_{1-\gamma}^{1} \overline{U}^b(b)\, db$, which is greater than 1/2, by the above calculation.

By identical arguments, each seller's expected utility is also greater than 1/2.

But this means that when all agents' values are drawn from (F^b, F^s), the total expected surplus generated by the mechanism must be more than $2N \cdot 1/2 = N$. Since it is never possible to generate a surplus of more than N, we have a contradiction.

$\square$

4 A consequentialist approach

The exposition so far has focused on incentives for truthful reporting. This follows a substantial literature that treats strategyproofness as a basic normative criterion for evaluating mechanisms (see [5] for a survey, and [4] for a succinct summary of several justifications). However, others [6, 9] have raised the criticism that truthfulness should not be an end in itself; rather, what matters is the outcome that occurs as a result of any manipulations. In particular, in the double auction environment, there is an unambiguous objective available to a planner with such a "consequentialist" philosophy — namely, the efficiency of the realized allocation of goods — and so it is especially natural to frame the design problem in terms of this objective.

Fortunately, it turns out that there is a close connection between our formulation of the efficiency-incentive tradeoff and an alternative formulation that focuses on outcome efficiency. To motivate the latter formulation, imagine a planner who wants to ensure an allocation within η of the first-best welfare, and who is uncertain not only about the distributions (F^b, F^s) but also about the agents' strategic behavior. Thus, the planner wants to ensure that no matter what manipulations the agents perform, welfare is always within η of the first-best (in expectation over realizations of the agents' types).

To describe the planner's problem, we must specify how she expects agents to manipulate. As sketched in Subsection 2.1, we presume there is a computational cost of at least ϵ to behaving strategically, so the planner is confident that agents will not manipulate if they cannot gain more than ϵ expected utility by doing so. What if they can gain more than ϵ? We could assume that agents will choose the manipulation that is optimal (with respect to their beliefs), but this would stray from our motivating notion of inexperienced, boundedly-rational agents. Instead we will take a more agnostic approach: agents may potentially make any misreport that would gain at least ϵ expected utility.[3]

We formalize this approach as follows. Given a mechanism M, a class $\mathcal{F}$ of distribution pairs, and a minimum manipulation cost ϵ, define the *manipulation set* for each possible buyer's valuation $b_i \in [0,1]$ as

$$\begin{aligned} W^b(b_i;\epsilon) = \{b_i\} \cup \\ \{\widehat{b}_i \mid E_{(F^b,F^s)}[U_i^b(\widehat{b}_i, P_{-i}^b, P^s|b_i)] - E_{(F^b,F^s)}[U_i^b(b_i, P_{-i}^b, P^s)] \geq \epsilon \\ \text{for some } (F^b,F^s) \in \mathcal{F}\}. \end{aligned}$$

(This is independent of i, by anonymity.) This set represents the set of all valuations that the planner believes a buyer might report, given that his true valuation is b_i. Note that we always include b_i: no matter what the mechanism is, we allow for the possibility that strategizing is so costly that the buyer just tells the truth. Similarly, for each seller's valuation s_i we define

$$\begin{aligned} W^s(s_i;\epsilon) = \{s_i\} \cup \\ \{\widehat{s}_i \mid E_{(F^b,F^s)}[U_i^s(P^b, \widehat{s}_i, P_{-i}^s|s_i)] - E_{(F^b,F^s)}[U_i^s(P^b, s_i, P_{-i}^s)] \geq \epsilon \\ \text{for some } (F^b,F^s) \in \mathcal{F}\}. \end{aligned}$$

From the planner's point of view, each buyer's true valuation and his report are drawn from a *joint* distribution H^b on $[0,1] \times [0,1]$, independently across buyers. We say that such a joint distribution H^b is *possible* if it places probability 1 on the set of pairs $(b,\widehat{b})$ such that $\widehat{b} \in W^b(b;\epsilon)$; and similarly for joint distributions H^s of sellers' valuations and reports.

The planner's measure of inefficiency is given by the worst case over all possible *joint* distributions H^b and H^s. For any profile P of true valuations and $\widehat{P}$ of reports, define the realized welfare $W^M(\widehat{P}|P)$ as $\sum_i b_i p_i^b(\widehat{P}) + \sum_i s_i(1-p_i^s(\widehat{P}))$. Then define the *consequentialist inefficiency* of the mechanism M (with minimum manipulation cost ϵ) as

$$\eta^c = \sup_{(H^b,H^s)} \left(E_{(H^b,H^s)}[W^{FB}(P) - W^M(\widehat{P}|P)]\right),$$

where the expectation is over true profiles P and reported profiles $\widehat{P}$ obtained by drawing each $(b_i,\widehat{b}_i) \sim H^b$ and each $(s_i,\widehat{s}_i) \sim H^s$ independently, and the supremum is over pairs

[3] In [8], we gave a positive model that effectively assumes the planner considers any misreport to be possible if she is not certain that the agents cannot gain more than ϵ. The model here is more refined.

such that

- H^b is possible for the buyers,
- H^s is possible for the sellers, and
- the marginals F^b of H^b and F^s of H^s on true valuations satisfy $(F^b, F^s) \in \mathcal{F}$.

The value of η^c of course depends on ϵ. The higher ϵ is, the smaller the manipulation sets are, the smaller the set of (H^b, H^s) over which the sup is taken, and so the smaller is consequentialist inefficiency. Note also that the consequentialist inefficiency η^c is always at least as large as the truthful inefficiency η.

This leads to our main definition: we say that a mechanism M has a (σ, η) *consequentialist tradeoff* on the class of distribution pairs $\mathcal{F}$ if, for any manipulation cost $\epsilon < \sigma$, the mechanism's consequentialist inefficiency on $\mathcal{F}$ is at least η. This expresses the tradeoff faced by a planner: she must either be willing to assume that agents have a manipulation cost at least σ, or accept an allocative inefficiency of at least η.

With these definitions behind us, we can proceed to convert our results into the consequentialist framework. Our earlier results were of the form

> for a given class of distribution pairs $\mathcal{F}$, every mechanism either has inefficiency greater than *[bound]* or susceptibility greater than *[bound]*.

We would now like to have results of the form

> for a given $\mathcal{F}$, every mechanism has a (*[bound]*,*[bound]*) consequentialist tradeoff.

To make this leap, we focus on misreports that are not too small. The intuition is as follows: Suppose a mechanism makes a buyer of value b willing to misreport as a value $\widehat{b}$, and $\widehat{b}$ is far from b — say $\widehat{b} < b$ for example. Then when all other agents report values in between b and $\widehat{b}$, the mechanism cannot distinguish whether the buyer reporting $\widehat{b}$ actually has value $\widehat{b}$ (in which case efficiency would imply that the buyer should not get the good) or actually has value b (in which case the buyer should get the good). So whatever allocation the mechanism specifies will be bounded away from efficiency in one of the two cases.

Once again, the intuition requires some elaboration because of our restriction to continuous distributions — misreports by just a single type, or by a finite set of types, have zero effect on expected efficiency. The technical apparatus needed to make the argument work is as follows.

We define a *quasi-misreport* for a buyer to be a triple $(b, \widehat{b}, \delta)$, where $\delta \in [0,1]$ and $b, \widehat{b} \in [\delta, 1-\delta]$, and $\widehat{b} \neq b$. The interpretation of a quasi-misreport is not just that buyers of type b are willing to misreport as $\widehat{b}$, but rather that a positive measure of types b' within δ of b are each incentivized to misreport by the amount $\widehat{b} - b$. A quasi-misreport for a seller is analogously a triple $(s, \widehat{s}, \delta)$.

Formally: we say that the mechanism M is *σ-susceptible* to the quasi-misreport $(b, \widehat{b}, \delta)$ of a buyer under $\mathcal{F}$, if the set

$$\{b'_i \in [b-\delta, b+\delta] \mid b'_i + (\widehat{b} - b) \in W^b(b'_i; \sigma)\}$$

has positive Lebesgue measure. We define σ-susceptibility to quasi-misreports of a seller analogously.

It is clear that if a mechanism is σ-susceptible to any quasi-misreport, then it has susceptibility at least σ. Thus we can think of susceptibility to a particular set of quasi-misreports as a strengthening of susceptibility. This strengthening ties in with consequentialist inefficiency via the following lemma:

Lemma 5 *Assume $N \geq 2$. Let $\mathcal{F}$ be a set of distribution pairs with $\mathcal{F}_1 \subseteq \mathcal{F}$.*

If M is σ-susceptible to the buyer's quasi-misreport $(b, \widehat{b}, \delta)$, and

$$\eta < \frac{b - \widehat{b} - 4\delta}{64}, \tag{12}$$

then M has a (σ, η) consequentialist tradeoff over $\mathcal{F}$. Similarly, if M is σ-susceptible to the seller's quasi-misreport $(s, \widehat{s}, \delta)$, and

$$\eta < \frac{\widehat{s} - s - 4\delta}{64}, \tag{13}$$

then M has a (σ, η) consequentialist tradeoff over $\mathcal{F}$.

Proof: We give the proof for (12); the argument for (13) is essentially identical. Note (12) implies $b - \widehat{b} \geq 4\delta$.

Let R be the set of values $r \in [-\delta, \delta]$ such that a buyer of type $b + r$ can benefit by at least σ from misreporting as $\widehat{b} + r$, for some distribution pair in $\mathcal{F}$. Thus, for every $r \in R$, the manipulation set $W^b(b + r; \sigma)$ contains $\widehat{b} + r$, and R has Lebesgue measure $\mu > 0$.

Define density functions f, g as follows:

$$f(x) = \begin{cases} 2/\delta, & \frac{b+\widehat{b}}{2} - \delta < x < \frac{b+\widehat{b}}{2} + \delta \\ 0 & \text{otherwise;} \end{cases}$$

$$g(x) = \begin{cases} 1/\mu, & x = b + r \text{ for some } r \in R \\ 0 & \text{otherwise.} \end{cases}$$

Define the density $h(x) = \left(1 - \frac{1}{N}\right) f(x) + \left(\frac{1}{N}\right) g(x)$. Let F, G, H be the associated distributions.

Drawing an agent's value from H is equivalent to designating the agent as "F-type" or "G-type" with probabilities $1 - 1/N$ or $1/N$, respectively, then drawing a valuation from F or G accordingly.

Certainly $(H, H) \in \mathcal{F}$. Suppose all agents' values are drawn independently from H, and that the agents report as follows: any G-type buyer misreports by $\widehat{b} - b$ (so if his true value is $b + r$, he reports $\widehat{b} + r$); all other agents report truthfully.

Let E denote the event that there is exactly one G-type buyer, and the $2N - 1$ other agents are all F-type. As in (4), we have

$$Pr(E) \geq \frac{1}{16}. \tag{14}$$

Conditional on E, let π be the probability that the G-type buyer ends up with a good under the mechanism. Notice that in event E, the first-best always requires this buyer to receive a good; when he does not, the resulting efficiency loss is at least as large as the difference between his value and the next-highest value, which is at least

$$(b - \delta) - \left(\frac{b + \widehat{b}}{2} + \delta\right) = \frac{b - \widehat{b}}{2} - 2\delta.$$

Therefore, the consequentialist inefficiency of the mechanism (with minimum manipulation cost σ) satisfies the lower bound

$$\eta^c \geq Pr(E) \cdot (1 - \pi) \cdot \left(\frac{b - \widehat{b}}{2} - 2\delta\right) \geq \frac{\frac{b-\widehat{b}}{2} - 2\delta}{16} \cdot (1 - \pi). \tag{15}$$

On the other hand, let $\widehat{g}$ be the density defined by

$$\widehat{g}(x) = \begin{cases} 1/\mu, & x = \widehat{b} + r \text{ for some } r \in R \\ 0 & \text{otherwise.} \end{cases}$$

and $\widehat{h}(x) = \left(1 - \frac{1}{N}\right) f(x) + \left(\frac{1}{N}\right) \widehat{g}(x)$. Define $\widehat{G}, \widehat{H}$ the distributions associated with $\widehat{g}, \widehat{h}$. Note that $\widehat{G}$ represents the distribution of *reports* by a G-type buyer in the previous scenario who misreports his value.

Suppose now that all agents' values are drawn from $\widehat{H}$, instead of H, and that all agents report truthfully. We can label agents as F-type or $\widehat{G}$-type, as before. Let $\widehat{E}$ denote the event that there is one $\widehat{G}$-type buyer and all other agents are F-type. We then have $Pr(\widehat{E}) \geq 1/16$ once again. Moreover, the distribution of profiles conditional on $\widehat{E}$, when values are drawn from $\widehat{H}$, is exactly the same as the distribution of *reported* profiles conditional on E, when values were drawn from H and when G-type buyers were misreporting. Consequently, conditional on $\widehat{E}$, the probability that the $\widehat{G}$-type buyer receives a good is again π.

However, conditional on $\widehat{E}$, the first-best requires that the $\widehat{G}$-type buyer never receive a good; and when he does receive one, the efficiency loss is at least the difference between his value and the next higher value, which is at least

$$\left(\frac{b + \widehat{b}}{2} - \delta\right) - \left(\widehat{b} + \delta\right) = \frac{b - \widehat{b}}{2} - 2\delta.$$

So we have

$$\eta \geq Pr(\widehat{E}) \cdot \pi \cdot \left(\frac{b - \widehat{b}}{2} - 2\delta\right) \geq \frac{\frac{b - \widehat{b}}{2} - 2\delta}{16} \cdot \pi. \tag{16}$$

Adding (15) and (16), and dividing by 2 gives $\eta^c \geq (b - \widehat{b} - 4\delta)/64$. Combining with (12), we see that the mechanism has a (σ, η) consequentialist tradeoff, as claimed.

□

We can now use Lemma 5 to restate our main results from Section 3 in terms of consequentialist tradeoffs. The following theorem extends Theorem 3:

Theorem 6 *There exists a positive constant c such that every possible mechanism has a $(c/(\sqrt{N}\log N), c/(\sqrt{N}\log N))$ consequentialist tradeoff on $\mathcal{F}_1$.*

Proof: Suppose not: some mechanism M has consequentialist inefficiency η^c less than $c/(\sqrt{N}\log N)$ for manipulation cost $\sigma < c/(\sqrt{N}\log N)$. We repeat exactly the steps of the proof of Theorem 3. Since $\eta < \eta^c$, the only assumption from that theorem that is no longer present was the assumption that each agent can gain at most σ by misreporting. That assumption was used only once in the original proof — in Step 2, in the line "a buyer of type $b_j + r$ (for any j) cannot benefit by more than σ from misreporting as type $b_{j+1} + r$" (and the analogous argument for sellers). This line now requires elaboration. In particular, it must be reformulated in terms of quasi-misreports.

We claim that for each j, the mechanism M cannot be σ-susceptible to the quasi-misreport $\left(b_j, b_{j+1}, \frac{\gamma}{2}\right)$. For suppose otherwise. Then by Lemma 5, we have

$$\eta \geq \frac{b_j - b_{j+1} - 2\gamma}{64} > \frac{b_j - b_{j+1}}{70}.$$

Now, the ratio ρ satisfies the lower bound

$$\rho \geq \frac{3}{4}(20\sqrt{N})^{1/K} \geq \frac{3}{4}(\sqrt{N})^{1/\log N} = \frac{3}{4}e^{1/2} > \frac{6}{5}, \tag{17}$$

which gives us

$$b_j - b_{j+1} = (\rho - 1)b_{j+1} \geq \frac{b_{j+1}}{5} \geq \frac{1}{100\sqrt{N}}$$

and therefore

$$\eta \geq \frac{1/100\sqrt{N}}{70} = \frac{1}{7000\sqrt{N}}.$$

Since $\eta < c/(\sqrt{N}\log N)$, we obtain $c \geq 1/7000$ (as long as $N \geq 3$), contradicting our assumption at the beginning of the proof.

Thus, M is not susceptible to the quasi-misreport $\left(b_j, b_{j+1}, \frac{\gamma}{2}\right)$. So it remains true that for *almost* all $r \in [-\gamma/2, \gamma/2]$, a buyer of type $b_j + r$ (for any given j) cannot benefit by more than σ from misreporting as type $b_{j+1} + r$. Hence, for almost all r, this holds for all j simultaneously.

For each such r, the argument leading to (7) remains valid. Then (8) continues to hold as well, since that inequality is derived by integrating over $r \in [-\gamma/2, \gamma/2]$ (and the integrand is bounded). Thus, the conclusion of Step 2 on the average utility of each buyer still applies. An entirely analogous argument shows that we also still have the same lower bound on average utility for the sellers.

From there, the rest of the argument for Theorem 3 leads to the same contradiction as before. $\square$

Similarly, the following result extends Proposition 4 to the consequentialist framework:

Proposition 7 *There exists a positive constant c such that every mechanism has a (c, c) consequentialist tradeoff on $\mathcal{F}_\infty$.*

Proof: We prove the proposition with $c = 1/128$. Thus suppose for contradiction that some mechanism has consequentialist inefficiency $\eta^c < 1/128$ with $\sigma < 1/128$. Again, we repeat line-for-line the proof of Proposition 4, making a change analogous to the one we applied to prove Theorem 6.

Specifically, the only line in the proof of Proposition 4 that needs to be changed is the assertion "for any $r \in [0, \gamma]$, a buyer of value $1 - r$ cannot benefit by more than σ by misreporting as value $1/4 - r$." Instead, this line now only holds for almost all $r \in [0, \gamma]$; that is, the mechanism is not σ-susceptible to the quasi-misreport $\left(1 - \frac{\gamma}{2}, \frac{1}{4} - \frac{\gamma}{2}, \frac{\gamma}{2}\right)$. Proof: if it were σ-susceptible, then by Lemma 5, we would have

$$\eta \geq \frac{3/4 - 2\gamma}{64} > \frac{1}{128},$$

contrary to assumption. (An analogous change would be made in the argument for sellers.)

Again, the fact that misreports are prevented for almost all $r \in [0, \gamma]$ rather than all r is immaterial, since the proof of Proposition 4 then proceeds by integrating over r. The rest of that proof then carries through, and we reach the same contradiction.

$\square$

To summarize this section: although our main results were originally expressed in terms of the tradeoff between incentives for strategic manipulation and efficiency under truth-telling, they can be easily rephrased in terms of the tradeoff between costs of strategic behavior and efficiency under manipulation. The proofs carry over with only minor enhancements needed.

Before closing, we should mention that all of the above discussion has used only the allocation of goods as the relevant welfare criterion. In fact, with our assumption that agents face a cost to behaving strategically, it would arguably be appropriate to count this cost as a welfare loss whenever it is incurred. Of course, doing so would only strengthen our lower bounds on consequentialist inefficiency.

5 Onwards

In this paper, we have looked at the tradeoff between efficiency and incentives for strategic manipulation in large double auction mechanisms. In so doing, we have begun to fill a gap between two earlier literatures on large double auctions — one looking only at incentives for manipulation, and one looking at inefficiency in perfectly incentive-compatible mechanisms. By looking at the tradeoff, we have addressed the question of whether it would be possible to achieve much-improved convergence to full efficiency by making a small sacrifice in terms of incentives for truthful behavior. Our main result, Theorem 3, gives a negative answer to this question, by providing a near-optimal bound for the rate at which either the inefficiency or the susceptibility to manipulation of any mechanism can converge to zero as the size of the market becomes large. We have also reinterpreted the bound in terms of the severity of inefficiency that may result when agents actually do manipulate (Theorem 6).

There are several clear technical directions in which to extend this paper. One direction would be to strengthen Theorem 3 to give a sharp bound on the inefficiency-susceptibility tradeoff. Ideally, it should be possible to parameterize the curve to which the inefficiency-susceptibility frontier (depicted in Figure 1) converges as N becomes large, analogously to the deficit-inefficiency frontier parameterized by Tatur [31].

One might also wish to give analogous bounds for other classes of distribution pairs $\mathcal{F}$, besides those we have looked at. For example, one might consider the family of all pairs (F^b, F^s) given by continuous densities taking values in some interval $[\underline{p}, \overline{p}]$, where $0 < \underline{p} < \overline{p}$ are fixed; this would be more comparable with previous literature [13, 23, 28, 31].

The present paper fits into the program advanced in [8], which argues that it can be useful to quantify incentives for strategic behavior in mechanisms, and that a natural approach to doing so — defining a mechanism's susceptibility to manipulation as the maximum expected utility an agent could gain by manipulating — is analytically tractable. By looking at incentives in this way, rather than treating incentive constraints as rigid, we open up a new quantitative dimension to mechanism design. Understanding this dimension may be useful in designing and evaluating mechanisms for practical use.

A Omitted proofs

We begin by introducing some asymptotic notation used in the proofs. We follow the conventions of [8] and keep explicit track of constant factors. Specifically, for functions $F(N), G(N)$, we write $F(N) \sim G(N)$ to mean that $F(N)/G(N) \to 1$ as $N \to \infty$, and $F(N) \lesssim G(N)$ to mean $\limsup_{N\to\infty} F(N)/G(N) \leq 1$.

Now, we prove a technical result, Lemma 9, that is used in the proofs of Propositions 1 and 2. It provides a central-limit-theorem-style approximation on the probability of a given split between the number of high-value and the number of low-value agents.

We first need the following preliminary calculation:

Lemma 8 *Fix $0 < \kappa < 1$. Then*

$$\max_{0 \le J \le N} \binom{N}{J} \kappa^J (1-\kappa)^{N-J} \lesssim \frac{1}{\sqrt{2\pi\kappa(1-\kappa)N}}.$$

Proof: The maximum of the left-hand side over J is attained at $J = \lfloor (N+1)\kappa \rfloor$ (this can be proven by computing the ratio of its values over successive J). Now expand explicitly, use Stirling's approximation [1, eq. 6.1.38] for the factorials, and simplify. □

Lemma 9 *Let $0 < \kappa < 1$ and $\lambda \ge 1$ be given. There exist a constant c and an integer N_0 with the following property: For all $N > N_0$, all $K \le (1-\kappa)N$, and all $a, b \in [0,1]$ such that*

$$b \le \lambda a, \qquad 1 - b \le \lambda(1-a),$$

we have the inequality

$$\sum_{J=K}^{N} \binom{N}{J} \binom{N}{J-K} a^{N-J} b^{J-K} (1-a)^J (1-b)^{N-J+K} \le c \sqrt{\frac{\lambda}{\kappa N}}. \tag{18}$$

Proof: We consider three cases, depending on the values of a and b.

(i) Suppose that $a < \kappa/4\lambda$. Then $b < \kappa/4$. For every J we have either

$$\frac{\kappa}{2\lambda} \le \frac{N-J}{N} \tag{19}$$

or

$$\frac{\kappa}{2} \le \frac{J-K}{N}, \tag{20}$$

since otherwise adding would give $\frac{N-K}{N} < \kappa \frac{1+1/\lambda}{2} \le \kappa$, a contradiction.

If (19) holds then consider

$$\binom{N}{J} a^{N-J} (1-a)^J$$

which is log-concave in a, maximized at $a = (N-J)/N$. The constraint $a < \kappa/4\lambda$

then implies

$$\begin{aligned}
\binom{N}{J} a^{N-J}(1-a)^J &< \binom{N}{J}\left(\frac{\kappa}{4\lambda}\right)^{N-J}\left(1-\frac{\kappa}{4\lambda}\right)^J \\
&\lesssim \frac{1}{\sqrt{2\pi N\left(\frac{\kappa}{4\lambda}\right)\left(1-\frac{\kappa}{4\lambda}\right)}}
\end{aligned}$$

using Lemma 8.

If (20) holds then consider

$$\binom{N}{J-K} b^{J-K}(1-b)^{N-J+K},$$

which is log-concave in b, maximized at $b = (J-K)/N$. The constraint $b < \kappa/4$ implies

$$\begin{aligned}
\binom{N}{J-K} b^{J-K}(1-b)^{N-J+K} &< \binom{N}{J-K}\left(\frac{\kappa}{4}\right)^{J-K}\left(1-\frac{\kappa}{4}\right)^{N-J+K} \\
&\lesssim \frac{1}{\sqrt{2\pi N\left(\frac{\kappa}{4}\right)\left(1-\frac{\kappa}{4}\right)}}
\end{aligned}$$

again by Lemma 8.

So there is an absolute constant c such that, for every J, one of the two factors

$$\binom{N}{J} a^{N-J}(1-a)^J, \qquad \binom{N}{J-K} b^{J-K}(1-b)^{N-J+K}$$

is at most $c\sqrt{\lambda/\kappa N}$ (as long as N is large enough). Then the sum in (18) is at most

$$c\sqrt{\frac{\lambda}{\kappa N}}\left[\sum_{J=K}^{N}\binom{N}{J-K} b^{J-K}(1-b)^{N-J+K} + \sum_{J=K}^{N}\binom{N}{J} a^{N-J}(1-a)^J\right]$$

$$\leq c\sqrt{\frac{\lambda}{\kappa N}}\left[(b+(1-b))^N + (a+(1-a))^N\right] = 2c\sqrt{\frac{\lambda}{\kappa N}}.$$

(ii) Suppose that $1-a < \kappa/4\lambda$. Then $1-b < \kappa/4$. Here the analysis is quite similar to case (i): For every J we have either

$$\frac{1}{2} \leq \frac{J}{N} \tag{21}$$

or

$$\frac{1}{2} \leq \frac{N-J+K}{N}. \tag{22}$$

If (21) holds then

$$\begin{aligned}\binom{N}{J} a^{N-J}(1-a)^J &< \binom{N}{J}\left(1-\frac{\kappa}{4\lambda}\right)^{N-J}\left(\frac{\kappa}{4\lambda}\right)^J \\ &\lesssim \frac{1}{\sqrt{2\pi N\left(1-\frac{\kappa}{4\lambda}\right)\left(\frac{\kappa}{4\lambda}\right)}}\end{aligned}$$

and if (22) holds then

$$\begin{aligned}\binom{N}{J-K} b^{J-K}(1-b)^{N-J+K} &< \binom{N}{J-K}\left(1-\frac{\kappa}{4}\right)^{J-K}\left(\frac{\kappa}{4}\right)^{N-J+K} \\ &\lesssim \frac{1}{\sqrt{2\pi N\left(1-\frac{\kappa}{4}\right)\left(\frac{\kappa}{4}\right)}}.\end{aligned}$$

This case is completed exactly as in the previous case.

(iii) The remaining possibility is $\kappa/4\lambda \leq a \leq 1-\kappa/4\lambda$. In this case, we hold a fixed and let N and J vary. We use Lemma 8, which gives

$$\max_J \binom{N}{J} a^{N-J}(1-a)^J \lesssim \frac{1}{\sqrt{2\pi N a(1-a)}} \leq c\sqrt{\frac{\lambda}{\kappa N}}$$

for an appropriate constant c. Then the sum in (18) is at most

$$c\sqrt{\frac{\lambda}{\kappa N}} \sum_{J=K}^{N} \binom{N}{J-K} b^{J-K}(1-b)^{N-J+K} \leq c\sqrt{\frac{\lambda}{\kappa N}}[(b+(1-b))^N] = c\sqrt{\frac{\lambda}{\kappa N}}.$$

□

Proof of Proposition 1: We will show the following stronger result: there is an absolute constant c such that the expected value of the least valuable trade, under any distribution pair $(F^b, F^s) \in \mathcal{F}_\lambda$, is at most $c\lambda^{5/2}N^{-1/2}$, as long as N is sufficiently large relative to λ. Denote this expected value by $\zeta(F^b, F^s)$.

First, fix any N and any $(F^b, F^s) \in \mathcal{F}$. For each $x \in [0,1]$, let $H(x)$ denote the probability that $s_{(k)} < x < b_{(k)}$, where $s_{(k)}, b_{(k)}$ denote the values involved in the lowest-value trade as in Subsection 2.3. Conditional on the realized profile, the value of this lowest-value trade, $b_{(k)} - s_{(k)}$, equals the probability that $s_{(k)} < x < b_{(k)}$ when x is drawn uniformly from $[0,1]$. Hence, the unconditional expected value of $b_{(k)} - s_{(k)}$ is just the expected value of $H(x)$, over $x \sim U[0,1]$. That is,

$$\zeta(F^b, F^s) = E[b_{(k)} - s_{(k)}] = \int_0^1 H(x)\, dx.$$

So it suffices to show that $\max_{x \in [0,1]} H(x)$ is bounded above by $c\lambda^{5/2}N^{-1/2}$.

Thus, fix $x^* \in [0,1]$. Call a valuation *high* if it is in $[x^*, 1]$ and *low* if it is in $[0, x^*)$. Notice that $b_{(k)}$ is the lowest buyer's value among the top N values, and $s_{(k)}$ is the highest seller's value among the bottom N values. Therefore, $s_{(k)} < x^* < b_{(k)}$ if and only if all buyers among the top N values are high and all sellers among the bottom N values are low. Call this event E^*. Thus, $H(x^*) = Pr(E^*)$.

To bound the probability of E^*, we define the following events:

- E_K, for each integer $K = -N, -N+1, \ldots, N$, is the event that there are exactly $N + K$ high values.

- E'_K, for $K = 0, \ldots, N$, is the event that E_K happens and the $(N+1)$th, $\ldots$, $(N + K)$th highest values are all buyer values.

- E'_K, for $K = -N, \ldots, -1$, is the event that E_K happens and the Nth, $(N-1)$th, $\ldots$, $(N + K + 1)$th highest values are all seller values.

Note that E^* is contained in the union of the E'_K.

We claim that for $|K| \leq N/2$, $Pr(E_K) \leq c\lambda^{1/2}N^{-1/2}$, where c is an absolute constant (as long as N is large enough). Indeed, if we let J denote the number of high buyer values, we can sum over possible realizations of J to obtain (when $K \geq 0$) the equality

$$Pr(E_K) = \sum_{J=K}^{N} \binom{N}{J} \binom{N}{N+K-J} F^b(\gamma)^{N-J} F^s(\gamma)^{J-K} (1 - F^b(\gamma))^J (1 - F^s(\gamma))^{N+K-J}. \tag{23}$$

A direct application of Lemma 9, with $\kappa = 1/2$, then implies that $Pr(E_K) \leq c\lambda^{1/2}N^{-1/2}$ as claimed. The argument for the case $K < 0$ is identical.

Next, we claim that

$$Pr(E'_K | E_K) \leq \left(1 + \frac{1}{2\lambda^2}\right)^{-|K|}, \qquad \text{for } |K| \leq \frac{N}{2}. \tag{24}$$

To show this, we argue in terms of the joint density of the $2N$ values (b_i, s_i). We will again assume $K \geq 0$; the argument for $K < 0$ is identical.

For any weakly decreasing sequence of values $v = (v_{(1)} \geq \cdots \geq v_{(2N)})$ and any sequence of labels $t = (t_{(1)}, \ldots, t_{(2N)})$ with each $t_{(i)} \in \{\mathtt{b}, \mathtt{s}\}$, let

$$Q(v, t) = \prod_{i:\ t_{(i)}=\mathtt{b}} f^b(v_{(i)}) \cdot \prod_{i:\ t_{(i)}=\mathtt{s}} f^s(v_{(i)}).$$

If the buyers' and sellers' values are drawn independently from F^b and F^s, then the probability density of a given profile P of values is exactly $Q(v, t)$, where v consists of the values in P sorted in decreasing order, and $t_{(i)} = \mathtt{b}$ if the value $v_{(i)}$ belongs to a buyer and $\mathtt{s}$ if a seller. For any set T of label sequences t, let $Q(v, T) = \sum_{t \in T} Q(v, t)$. For $J = 0, \ldots, K$, let T_J be the set of label sequences consisting of N $\mathtt{b}$'s and N $\mathtt{s}$'s, such that exactly J of the labels $t_{(N+1)}, \ldots, t_{(N+K)}$ are equal to $\mathtt{s}$; and let $T_\cup = \cup_{J=0}^{K} T_J$, the set of all label sequences consisting of N $\mathtt{b}$'s and N $\mathtt{s}$'s.

Let V_K be the set of value sequences consisting of $N + K$ high values and $N - K$ low values. Then

$$Pr(E_K) = (N!)^2 \int_{V_K} Q(v, T_\cup)\, dv. \tag{25}$$

(The $(N!)^2$ factor comes from the fact that each sequence v of distinct values and label sequence t distinguishing the buyer values from the seller values should be counted multiple times, once for each of the $N!$ possible assignments of buyer identities to buyer values and $N!$ assignments of seller identities to seller values.) Similarly

$$Pr(E'_K) = (N!)^2 \int_{V_K} Q(v, T_0)\, dv. \tag{26}$$

On the other hand, for any fixed v and any fixed $J \in \{0, \ldots, K-1\}$, we can relate $Q(v, T_J)$ with $Q(v, T_{J+1})$ as follows. Call an element $t_J \in T_J$ and $t_{J+1} \in T_{J+1}$ *connected* if t_{J+1} is obtained from t_J by switching some $t_{(i)}$ from $\mathtt{b}$ to $\mathtt{s}$, where $i \in \{N+1, \ldots, N+K\}$, and switching some $t_{(j)}$ from $\mathtt{s}$ to $\mathtt{b}$, where $j \notin \{N+1, \ldots, N+K\}$. Each element of T_J is connected to exactly $(K-J)(N-J)$ elements of T_{J+1}, and each element of T_{J+1} is connected to exactly $(J+1)(N-K+J+1)$ elements of T_J. Moreover, if t_{J+1} is connected to t_J, then $Q(v, t_J) \leq \lambda^2 Q(v, t_{J+1})$, since the ratio between f^b and f^s is always bounded by λ. Summing over all connected pairs, we have

$$(K-J)(N-J)\sum_{t_J\in T_J} Q(v,t_J) \le (J+1)(N-K+J+1)\sum_{t_{J+1}\in T_{J+1}} \lambda^2 Q(v,t_{J+1})$$

from which

$$Q(v,T_{J+1}) \ge \frac{(K-J)(N-J)}{(J+1)(N-K+J+1)\lambda^2} Q(v,T_J).$$

Since $N \ge 2K$ and $J \le K-1$ this gives

$$Q(v,T_{J+1}) \ge \frac{K-J}{J+1}\cdot\frac{1}{2\lambda^2}Q(v,T_J) = \frac{\binom{K}{J+1}}{\binom{K}{J}}\cdot\frac{1}{2\lambda^2}Q(v,T_J).$$

Now by induction we have

$$Q(v,T_J) \ge \binom{K}{J}\cdot\left(\frac{1}{2\lambda^2}\right)^J Q(v,T_0)$$

for all J. Summing gives

$$Q(v,T_\cup) = \sum_{J=0}^{K} Q(v,T_J) \ge \left(1+\frac{1}{2\lambda^2}\right)^K Q(v,T_0).$$

Combining with (25) and (26) gives

$$Pr(E_K) \ge \left(1+\frac{1}{2\lambda^2}\right)^K Pr(E'_K).$$

This is exactly (24) for $K \ge 0$. The $K < 0$ case is identical.

In addition, if $K > N/2$, then any draw in E'_K requires the $(N+1)$th, $\ldots$, $\lfloor 3N/2\rfloor$th highest values all to be buyer values; so an identical argument gives

$$Pr(E'_K|E_K) \le \left(1+\frac{1}{2\lambda^2}\right)^{-\lfloor N/2\rfloor}$$

for $K > N/2$. And by the same argument, this conclusion also holds when $K < -N/2$.

We conclude that

$$\begin{aligned}
Pr(E^*) &\le \sum_{K=-N}^{N} Pr(E'_K) \\
&\le \sum_{K=-\lfloor N/2 \rfloor}^{\lfloor N/2 \rfloor} c\lambda^{1/2} N^{-1/2} \cdot \left(1 + \frac{1}{2\lambda^2}\right)^{-|K|} + \sum_{\substack{-N \le K \le N \\ |K| > N/2}} \left(1 + \frac{1}{2\lambda^2}\right)^{-\lfloor N/2 \rfloor} \\
&\le 2 \sum_{K=0}^{\infty} \left(1 + \frac{1}{2\lambda^2}\right)^{-K} \cdot c\lambda^{1/2} N^{-1/2} + (N+2)\left(1 + \frac{1}{2\lambda^2}\right)^{-\lfloor N/2 \rfloor} \\
&\le 7c\lambda^{5/2} N^{-1/2}.
\end{aligned}$$

The last inequality holds because $\sum_{K=0}^{\infty}(1 + 1/2\lambda^2)^{-K} = 2\lambda^2 + 1 \le 3\lambda^2$, and the final term $(N+2)(1+1/2\lambda^2)^{-\lfloor N/2 \rfloor}$ is exponentially decreasing in N, so is certainly at most $c\lambda^{5/2}N^{-1/2}$ when N is large enough.

Thus we have shown that there is an absolute constant c for which $H(x^*) = Pr(E^*) \le c\lambda^{5/2}N^{-1/2}$ when N is large enough. Moreover, at no step in the proof did we use the specific value of x^* or the distribution $(F^b, F^s) \in \mathcal{F}$; therefore the constant c and the threshold for N are independent of these choices. We conclude that $\sup_{(F^b,F^s)\in\mathcal{F}} \zeta(F^b, F^s) \le c\lambda^{5/2}N^{-1/2}$, which is what we wanted.

□

Finally, we prove a simple central-limit-theorem approximation used in the proof of Theorem 3.

Lemma 10 *If N is sufficiently large, then*

$$\sum_{K=0}^{N-\lfloor\sqrt{N}/4\rfloor} \binom{2N}{K}\left(\frac{1}{2}\right)^{2N} \ge \frac{1}{4}.$$

Proof: From Stirling's approximation, we have

$$\binom{2N}{K}\left(\frac{1}{2}\right)^{2N} \le \binom{2N}{N}\left(\frac{1}{2}\right)^{2N} \lesssim \sqrt{\frac{2}{\pi N}}$$

and in particular

$$\binom{2N}{K}\left(\frac{1}{2}\right)^{2N} < \frac{1}{\sqrt{N}}$$

for all K, as long as N is large enough. Then, we have

$$\sum_{K=0}^{N-\lfloor\sqrt{N}/4\rfloor} \binom{2N}{K}\left(\frac{1}{2}\right)^{2N} \geq \sum_{K=0}^{N} \binom{2N}{K}\left(\frac{1}{2}\right)^{2N} - \lfloor\sqrt{N}/4\rfloor\frac{1}{\sqrt{N}} \geq \frac{1}{4}.$$

$\square$

References

[1] Milton Abramowitz and Irene A. Stegun (1972), *Handbook of Mathematical Functions* (Washington: U.S. Government Printing Office), tenth printing.

[2] George A. Akerlof (1970), "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics* 84 (3), 488-500.

[3] Itai Ashlagi, Mark Braverman, and Avinatan Hassidim (2011), "Matching with Couples Revisited," extended abstract in *Proceedings of the 12th ACM Conference on Electronic Commerce* (EC-11), 335.

[4] Eduardo Azevedo and Eric Budish, "Strategyproofness in the Large as a Desideratum for Market Design," unpublished paper, Harvard University.

[5] Salvador Barberà (2001), "An Introduction to Strategy-Proof Social Choice Functions," *Social Choice and Welfare* 18 (4), 619-653.

[6] Dirk Bergemann and Stephen Morris (2005), "Robust Mechanism Design," *Econometrica* 73 (6), 1771-1813.

[7] Eleanor Birrell and Rafael Pass (2011), "Approximately Strategy-Proof Voting," in *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*, 67-72.

[8] Gabriel Carroll (2011), "A Quantitative Approach to Incentives: Application to Voting Rules," unpublished paper, MIT.

[9] Kim-Sau Chung and J. C. Ely (2007), "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies* 74 (2), 447-476.

[10] Edward H. Clarke (1971), "Multipart Pricing of Public Goods," *Public Choice* 11 (1), 17-33.

[11] Olivier Compte and Philippe Jehiel (2009), "Veto Constraint in Mechanism Design: Inefficiency with Correlated Types," *American Economic Journal: Microeconomics* 1 (1), 182-206.

[12] Lars Ehlers, Hans Peters, and Ton Storcken (2004), "Threshold Strategy-Proofness: On Manipulability in Large Voting Problems," *Games and Economic Behavior* 49 (1), 103-116.

[13] Thomas A. Gresik and Mark A. Satterthwaite (1989), "The Rate at Which a Simple Market Converges to Efficiency as the Number of Traders Increases: An Asymptotic Result for Optimal Trading Mechanisms," *Journal of Economic Theory* 48 (1), 304-332.

[14] Theodore Groves (1973), "Incentives in Teams," *Econometrica* 41 (4), 617-631.

[15] Nicole Immorlica and Mohammed Mahdian (2005), "Marriage, Honesty, and Stability," in *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (SODA-05), 53-62.

[16] Matthew O. Jackson (1992), "Incentive Compatibility and Competitive Allocations," *Economics Letters* 40 (3), 299-302.

[17] Matthew O. Jackson and Alejandro M. Manelli (1997), "Approximately Competitive Equilibria in Large Finite Economies," *Journal of Economic Theory* 77 (2), 354-376.

[18] Fuhito Kojima and Mihai Manea (2010), "Incentives in the Probabilistic Serial Mechanism," *Journal of Economic Theory* 145 (1), 106-123.

[19] Fuhito Kojima and Parag A. Pathak (2009), "Incentives and Stability in Large Two-Sided Matching Markets," *American Economic Review* 99 (3), 608-627.

[20] Anshul Kothari, David C. Parkes, and Subhash Suri (2005), "Approximately-Strategyproof and Tractable Multiunit Auctions," *Decision Support Systems* 39 (1), 105-121.

[21] Hitoshi Matsushima (2008), "Behavioral Aspects of Implementation Theory," *Economics Letters* 100 (1), 161-164.

[22] Hitoshi Matsushima (2008), "Role of Honesty in Full Implementation," *Journal of Economic Theory* 139 (1), 353-359.

[23] R. Preston McAfee (1992), "A Dominant Strategy Double Auction," *Journal of Economic Theory* 56 (2), 434-450.

[24] Frank McSherry and Kunal Talwar (2007), "Mechanism Design via Differential Privacy," in *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science* (FOCS-07), 94-103.

[25] Roger B. Myerson and Mark Satterthwaite (1983), "Efficient Mechanisms for Bilateral Trading," *Journal of Economic Theory* 29 (2), 265-281.

[26] Philip J. Reny and Motty Perry (2006), "Toward a Strategic Foundation for Rational Expectations Equilibrium," *Econometrica* 74 (5), 1231-1269.

[27] Donald John Roberts and Andrew Postlewaite (1976), "The Incentives for Price-Taking Behavior in Large Exchange Economies," *Econometrica* 44 (1), 115-127.

[28] Aldo Rustichini, Mark A. Satterthwaite, and Steven R. Williams (1994), "Convergence to Efficiency in a Simple Market with Incomplete Information," *Econometrica* 62 (5), 1041-1063.

[29] Mark A. Satterthwaite and Steven R. Williams (2002), "The Optimality of a Simple Market Mechanism," *Econometrica* 70 (5), 1841-1863.

[30] James Schummer (2004), "Almost-Dominant Strategy Implementation: Exchange Economies," *Games and Economic Behavior* 48 (1), 154-170.

[31] Tymon Tatur (2005), "On the Tradeoff between Deficit and Inefficiency and the Double Auction with a Fixed Transaction Fee," *Econometrica* 73 (2), 517-570.

[32] Steven R. Williams (1999), "A Characterization of Efficient, Bayesian Incentive Compatible Mechanisms," *Economic Theory* 14 (1), 155-180.