Efficient Mechanisms with Risky Participation

Ruggiero Cavallo
Yahoo! Research
111 West 40th Street
New York, NY 10018
cavallo@yahoo-inc.com

Abstract

There is a fundamental incompatibility between efficiency, interim individual rationality, and budget-balance in mechanism design, even for extremely simple settings. Yet it *is* possible to specify efficient mechanisms that satisfy participation and budget-balance constraints in expectation, prior to types being realized. We do so here, in fact deriving mechanisms that are individually rational for each agent even ex post of *other* agents' type realizations. However, participation must still bear some risk of loss. For agents that are risk neutral, we show how the center can extract the entire surplus in expectation, or alternatively provide an equal expected share of the surplus for each participant, without violating dominant strategy incentive compatibility, efficiency, or ex ante budget-balance. We compare these solutions to a third efficient mechanism we design explicitly to address risk aversion in trade settings: payments are defined to minimize the odds of loss, satisfying ex ante participation constraints for agents with attitudes toward risk ranging from neutrality to high loss-aversion.

1 Introduction

We address a general problem of efficient decision-making amongst self-interested agents when commitment can be established prior to values being learned. Let us start with an example: Two software companies decide to collaborate on building a new web technology in advance of the upcoming summer Olympic games two years in the future. It cannot be determined in advance of the games what the most advantageous way of using the technology will be: whether both companies should provide it, whether just one should get full control, whether it should be sold off, etc. Eventually each company (agent) will privately learn its own value for each of the possible choices and a decision will be made.

Can a budget-balanced payment mechanism be designed—enforced by a third-party ("the center")[1] that shares the agents' expectations about how their own values will be realized—that achieves a social-welfare maximizing outcome in dominant strategies while, at the time the business union is formed, establishing that each agent and the center expects to gain from joining the venture? Furthermore, can we identify a mechanism that strikes an effective balance between the likely benefit to each agent and to the center, so that even those who are significantly loss-averse choose to participate?

Mechanism design uses monetary payments as a tool to achieve desirable outcomes in decision settings with self-interested agents, such as the above. Often the goal is to maximize social welfare (efficiency). In addition to (and in support of) efficiency, the properties of individual rationality (IR) (no agent is made worse off from participating) and no-deficit (net transfers from the mechanism to the agents are non-positive) are typically essential. Consider, for instance, bilateral trade, where a good is initially held by one agent and there is another who potentially values the good higher than the first. Ideally, the good would change hands if and only if the second agent's value is higher, and additionally both agents would end up at least as well off from having participated as not, but without requiring a subsidy from the mechanism. Unfortunately this ideal is unreachable, as demonstrated by the Myerson-Satterthwaite impossibility theorem [Myerson and Satterthwaite, 1983].

In the face of this theorem there has been significant work in designing mechanisms that do not achieve efficiency but maintain interim[2] IR and no-deficit (see, e.g., [Gresik and Satterthwaite, 1989; Yoon, 2008; Rustichini *et al.*, 1994; Tatur, 2005]). Although making sacrifices in social welfare is one natural response, another approach—novel, to our knowledge, which we initiate here—is to maintain dominant strategy efficiency and instead go outside the scope of the impossibility result by moving the individual rationality demand from interim to the requirement that each agent expect to gain from participating ex ante of his *own* type realization (given the probability distribution over his type) even *ex post* of the other agents' type realizations. This notion is significantly weaker than fully ex post IR but significantly stronger than fully ex ante IR, as it removes any need for agents to form expectations or reason about the types of others. Harking back to the software company collaboration example, be-

[1] One can imagine the center as an agent that has engineered the process, e.g., the parent company of subsidiaries that are run independently but could yield net efficiency gains through collaboration.

[2] Ex post of your *own* type realization, ex ante of others'.

fore learning its value function, company A should expect to gain from joining the venture regardless of how company B's value function eventually turns out. We demonstrate the existence of dominant strategy efficient solutions that always satisfy this IR notion and are ex ante no-deficit, for arbitrary distributions over types.

But, granting that a participation decision can be forced prior to learning values, if the proposition agents face is participation in a "risky" mechanism that is expected to bring gains but may bring losses, agent attitudes toward risk become critical. There is abundant evidence that in economic settings people are not completely neutral towards risk, but rather often manifest significant loss-aversion (see, e.g., [DellaVigna, 2009]). Yet the prevailing assumption in mechanism design research is risk neutrality,[3] an unsettling mismatch with reality. In this paper, in addition to designing mechanisms that have desirable revenue or fairness properties for risk neutral agents, we will provide a mechanism that is robust to a wide range of attitudes towards loss, including significant loss-aversion.

In the mechanisms we propose, the payment for each agent is defined in a way that exploits valuation information reported by the others. This makes the mechanisms especially compelling (less risky) when types are *correlated*, i.e., when considering any agent i, knowing the types of agents other than i provides good information about i's type. And, interestingly, the literature on mechanism design in correlated types settings helps motivate our concern for agent attitudes towards risk. Following important results by Cremér and McLean [1988] and McAfee and Reny [1992],[4] D'Aspremont et al. [1993] demonstrated that when an apparently rather mild condition is satisfied, technically *all* social choice functions (not just efficient ones) can be implemented in Bayes-Nash equilibrium. The tension between these theoretical results and practice casts suspicion on Bayes-Nash equilibrium as a solution concept and the assumption that prior beliefs are common across all agents,[5] but also strongly highlights the implausibility of the risk neutrality assumption.

In the current paper we obtain *dominant strategy* solutions, which allows us to move away from the risk neutrality assumption by incorporating loss-aversion (in Section 4), and also to impose a much more modest assumption about prior beliefs: we assume only that each agent i and the center form the same conditional distribution over i's type given any profile of realized types for the others; this plays no incentive role for the agent's type reporting (it is required only for the IR property), and agents need not share any common beliefs with *other agents* whatsoever.

[3]Exceptions include work addressing revenue maximization in auctions where either the seller or the buyers are risk-averse (see, e.g., [Maskin and Riley, 1984; Sundararajan and Yan, 2010]).

[4]For a single-item auction setting with common priors over the joint value distribution, [Cremér and McLean, 1988] specify a dominant strategy efficient mechanism satisfying ex ante IR in which the seller obtains the entire surplus.

[5]The common prior assumption is highly controversial because it seems to map so poorly to how real individuals form beliefs, as it implies that disagreement is solely a result of asymmetric information (see, e.g., [Morris, 1995]).

2 Preliminaries

There is a set of of agents $I = \{1, \ldots, n\}$ and a set of outcomes O. An agent $i \in I$ has type $\theta_i \in \Theta_i$ and value function $v_i : \Theta_i \times O \rightarrow \Re$. The joint type space is $\Theta = \Theta_1 \times \ldots \times \Theta_n$; $\theta \in \Theta$ is a vector of agent types $\theta_1, \ldots, \theta_n$; $\theta_{-i} \in \Theta_{-i}$ is the vector of types excluding that of agent i. $v(\theta, o) = \sum_{i \in I} v_i(\theta_i, o)$ and $v_{-i}(\theta_{-i}, o) = \sum_{j \in I \setminus \{i\}} v_j(\theta_j, o)$. We assume that $\forall \theta \in \Theta$, $\exists o \in O$ such that $v(\theta, o) \geq 0$. A choice function $f : \Theta \rightarrow O$ maps a profile of agent types to an outcome. We use f^* to denote an efficient choice function, and $f^*(\theta_{-i})$ for an outcome that maximizes the value to agents other than i, given their types θ_{-i}. That is, $\forall \theta \in \Theta$: $f^*(\theta) \in \arg\max_{o \in O} v(\theta, o)$ and $f^*(\theta_{-i}) \in \arg\max_{o \in O} v_{-i}(\theta_{-i}, o)$. In order to elicit truthful reporting of private information, we will be doing mechanism design (see, e.g., [Jackson, 2000] for an introduction). A mechanism (f, T) elicits a report $\hat{\theta}_i$ of each agent's private type, executes outcome $f(\hat{\theta})$, and then implements payments $T = (T_1, \ldots, T_n)$, where $T_i : \Theta \rightarrow \Re$ is a monetary transfer function defining the payment *to* agent i. We assume quasilinear utility. In this paper a choice function f^* and a truthful mechanism will be the context under which all values are considered, so we often write $v_i(\theta)$ and $v(\theta)$ as shorthand for $v_i(\theta, f^*(\theta))$ and $v(\theta, f^*(\theta))$.

To analyze IR properties we are concerned with the "net value" or surplus obtained by agents compared with in the initial state (e.g., in a trade setting, the outcome where no goods are exchanged), so we normalize valuations accordingly. For instance, in a bilateral trade example where the initial-holder of the good has value 0.4 and the buyer has value 0.7, letting o_2 and o_1 be the outcomes where, respectively, trade does and does not occur, valuations will be as in Table 1.

	v_1	v_2
o_1	0	0
o_2	-0.4	0.7

Table 1: Normalized value in a bilateral trade example.

A mechanism (f, T) is *ex ante individually rational (IR)* if and only if each agent's a priori expected equilibrium utility is non-negative, i.e., $\forall i \in I$, $\mathbb{E}_{\tilde{\theta}}[v_i(\tilde{\theta}_i, f(\tilde{\theta})) + T_i(\tilde{\theta})] \geq 0$, where there is a prior distribution over types and $\tilde{\theta}$ is a random variable representing the profile of types that will be realized. The mechanisms we propose will satisfy the following somewhat stronger IR property for each agent i, where the expectation is taken using the *actual* reported types of agents other than i, considering only i's type as an unknown variable:

Definition 1 (ex ante* individually rational)**.** *A truthful mechanism (f, T) is ex ante* IR if and only if $\forall i \in I$, $\forall \theta_{-i} \in \Theta_{-i}$,* $\mathbb{E}_{\tilde{\theta}_i}[v_i(\tilde{\theta}_i, f(\tilde{\theta}_i, \theta_{-i})) + T_i(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] \geq 0.$

Ex ante no-deficit is satisfied if and only if $\sum_{i \in I} \mathbb{E}_{\tilde{\theta}}[T_i(\tilde{\theta})] \leq 0$, and ex post no-deficit if and only if $\sum_{i \in I} T_i(\theta) \leq 0$, for every $\theta \in \Theta$. The ex ante properties are defined *for risk-neutral agents*; in the evaluation section we will see how the proposed mechanisms fare when participants (including the center) are instead loss-averse.

	∅	∅	VCG	AGV	several	this paper
effic.	•	○	•	○		•
BB	•	•	•	•	○	○
IR	•	•			•	○

Table 2: Summary of efficiency, budget-balance, and individual rationality properties in mechanism design. The Myerson-Satterthwaite impossibility result is reflected in the first two columns. • denotes *dominant strategy* efficiency, *ex post* IR, or *ex post* no-deficit; ○ denotes Bayes-Nash efficiency, *ex ante* IR, or *ex ante* no-deficit. AGV achieves *strong* budget-balance and the mechanisms in this paper achieve a stronger-than-ex-ante IR property.

2.1 Related previous mechanisms

Given a natural broadness condition on the typespace, dominant strategy truthful and efficient mechanisms are completely characterized by the Groves class [Holmstrom, 1979], which choose efficient outcomes and implement transfer function $T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) + h_i(\theta_{-i})$ for each agent i, for arbitrary function $h_i : \Theta_{-i} \to \Re$. That is, each agent's payment equals the others' aggregate reported value plus or minus some quantity independent of the agent's report. The VCG mechanism [Vickrey, 1961; Clarke, 1971; Groves, 1973] is a special instance of this class that defines $T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - v_{-i}(\theta_{-i}, f^*(\theta_{-i}))$.

The most well-known previous mechanism that is not geared towards satisfying interim IR is the AGV (or "expected externality") mechanism of Arrow [1979] and d'Aspremont and Gerard-Varet [1979]. In AGV each agent i is payed the expected value others will obtain given i's actual reported type, considering only the probability distribution over others' types, and then charged a budget balancing term independent of his report. AGV is *not* a Groves mechanism because the payment is the *expected* value achieved by the others rather than their actual obtained value, so truthfulness is obtained only in Bayes-Nash equilibrium. Moreover, truthfulness breaks down if types are correlated. VCG is ex post no-deficit and AGV is *strongly* budget-balanced (no deficit, no revenue). One reason these mechanisms are remarkable is that, moreover, in settings where agents obtain non-negative value for every outcome, VCG is ex post individually rational and AGV is ex ante individually rational. But many important settings *do* admit negative values. For instance, in the bilateral trade example depicted in Table 1, the optimal outcome from agent 2's perspective is that in which he gets the good, but agent 1's value there is -0.4. Without making an assumption akin to no-negative-values, neither VCG nor AGV satisfy even *ex ante* IR in settings such as bilateral trade.

So these previous mechanisms leave something important to be desired, and more broadly, by [Myerson and Satterthwaite, 1983] we know that even in simple settings there is *no* interim IR and no-deficit mechanism that is efficient even in Bayes-Nash equilibrium (see Table 2 for a summary of these results). But that negative fact still leaves the possibility of mechanisms that are *ex ante** IR, and that's what we explore.

3 Mechanisms

In this section we present three distinct mechanisms that divide the surplus in different ways. The first is oriented towards deficit avoidance, the second towards equal division, and the third is optimized to minimize risk and achieve participation in asymmetric settings with loss-aversion. In Section 4 we will evaluate and compare their performance.

3.1 Extracting all surplus in expectation

In the first mechanism we propose, the transfer for each agent is set in a manner that—in expectation—leads to extraction of the entire expected surplus of the mechanism by the center; yet ex ante* individual rationality is maintained.

Mechanism 1 (surplus extracting)**.** *A mechanism* (f^*, T)*, where,* $\forall i \in I$ *and* $\theta \in \Theta$*:*

$$T_i(\theta) = v_{-i}(\theta) - \mathbb{E}_{\tilde{\theta}_i}[v(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] \tag{1}$$

The mechanism makes the incentive-aligning Groves payment to each agent i, and charges each i an amount equal to the expected social-surplus that will result, given the type reports of the other agents but considering i's type as an unknown (with the expectation based on the prior distribution over θ_i).[6] To illustrate the mechanism, consider the bilateral trade example of Table 1 where the seller (agent 1) has value 0.4 and the other agent 0.7, and assume that, a priori, the distribution over each agent's value for the good is uniform on $[0, 1]$ and independent of the other agent's (his value for not receiving the good is known to be 0). We can compute:

$$T_1(\theta) = v_2(\theta) - \mathbb{E}_{\tilde{\theta}_1}[v_2(\tilde{\theta}_1, \theta_2) \,|\, \theta_2] - \mathbb{E}_{\tilde{\theta}_1}[v_1(\tilde{\theta}_1, \theta_2) \,|\, \theta_2]$$
$$= 0.7 - (0.7 \cdot 0.7 + 0.3 \cdot 0) - \Big(\int_0^{0.7} -x \, dx + 0.3 \cdot 0 \Big)$$
$$= 0.7 - 0.49 + \frac{0.49}{2} = 0.455, \text{ and}$$
$$T_2(\theta) = v_1(\theta) - \mathbb{E}_{\tilde{\theta}_2}[v_1(\theta_1, \tilde{\theta}_2) \,|\, \theta_1] - \mathbb{E}_{\tilde{\theta}_2}[v_2(\theta_1, \tilde{\theta}_2) \,|\, \theta_1]$$
$$= -0.4 + 0.6 \cdot 0.4 - \int_{0.4}^{1} x \, dx = -0.58$$

Agent 1 obtains a final net utility (including payments) of $-0.4 + 0.455 = 0.055$; agent 2 obtains $0.7 - 0.58 = 0.12$, and the center obtains $0.58 - 0.455 = 0.125$. Thus in this example there is no deficit and both agents are better off from having participated. We will see (in Theorem 2) that in this kind of bilateral trade setting a deficit never results. In the general case, with no assumptions about the domain or distributions over agent types, we have the following.

Theorem 1. *For all distributions over types, Mechanism 1 is truthful and efficient in dominant strategies, ex ante* individually rational, and ex ante no-deficit.*

Proof. The fact that the mechanism is truthful and efficient in dominant strategies follows from the fact that it is

[6] Expanding out the shorthand, Mechanism 1 defines: $T_i(\theta) = v_{-i}(\theta_{-i}, f^*(\theta)) - \mathbb{E}_{\tilde{\theta}_i}[v((\tilde{\theta}_i, \theta_{-i}), f^*(\tilde{\theta}_i, \theta_{-i})) \,|\, \theta_{-i}]$. Note that θ_i does not appear in the second term.

a Groves mechanism—the efficient outcome according to agent reports is chosen, and each agent i's payment equals the other agents' reported aggregate value minus a quantity ($\mathbb{E}_{\tilde{\theta}_i}[v(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}]$) independent of his report (recall that $\tilde{\theta}_i$ is a random variable, *not* a reported type).

Now note that any agent i's expected net utility prior to realization of his type, but given arbitrary θ_{-i}, is:

$$\mathbb{E}_{\tilde{\theta}_i}[v_i(\tilde{\theta}_i, \theta_{-i}) + v_{-i}(\tilde{\theta}_i, \theta_{-i}) - v(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] = 0, \quad (2)$$

so ex ante* individual rationality is satisfied.

Finally, consider the expected (ex ante) revenue to the center in the dominant strategy truthtelling equilibrium. Given the types of agents other than i, but considering i's type realization as unknown, the expected payment to the center from i is: $\mathbb{E}_{\tilde{\theta}_i}[v(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] - \mathbb{E}_{\tilde{\theta}_i}[v_{-i}(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] = \mathbb{E}_{\tilde{\theta}_i}[v_i(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}]$. So the total ex ante expected equilibrium revenue is:

$$\sum_{i \in I} \mathbb{E}_{\tilde{\theta}_{-i}} \left(\mathbb{E}_{\tilde{\theta}_i}[v_i(\tilde{\theta}_i, \tilde{\theta}_{-i}) \,|\, \tilde{\theta}_{-i}] \right) = \sum_{i \in I} \mathbb{E}_{\tilde{\theta}}[v_i(\tilde{\theta})] = \mathbb{E}_{\tilde{\theta}}[v(\tilde{\theta})] \quad (3)$$

This quantity is greater than or equal to 0 by efficiency of the mechanism, and thus ex ante no deficit holds. □

The following two theorems[7] demonstrate that the no-deficit property is *ex post* in important natural settings of bilateral trade, single-item allocation more generally, and even less structured domains. A single-item trade setting is one in which only one agent stands to obtain negative value from any outcome (the seller, if trade occurs); a distribution over agent types satisfies the no negative expected externalities condition if: $\forall i \in I, \forall \theta_{-i} \in \Theta_{-i}, \mathbb{E}_{\tilde{\theta}_i}[v_i(\tilde{\theta}_i, f^*(\theta_{-i})) \,|\, \theta_{-i}] \geq 0$.

Theorem 2. *Mechanism 1 is ex post no-deficit for all single-item trade settings where the distribution over values for the item is symmetric and independent across agents.*

Theorem 3. *Mechanism 1 is ex post no-deficit for all distributions over types that meet the no negative expected externalities condition.*

This last theorem helps specifically illustrate the tradeoff between the VCG mechanism and Mechanism 1. Both mechanisms are always efficient in dominant strategies. In no negative externalities settings VCG is ex post no-deficit *and* ex post IR, whereas Mechanism 1 is ex post no-deficit and only *ex ante** IR. But in settings where the no negative externalities condition is not met, Mechanism 1 retains ex ante* IR and ex ante or ex post no-deficit, while VCG does not retain even ex ante IR.

3.2 Evenly sharing the expected surplus

One could say Mechanism 1 achieves all the properties we could legitimately hope for, given the Myerson-Satterthwaite theorem and taking efficiency as a hard constraint. We know that improving the IR property to ex post is unattainable without giving something else up. However, that coarse level of analysis obscures the fact that under Mechanism 1 agents will clearly not fare as well as they would under another mechanism with the same high-level properties. The expected revenue generated by *any* mechanism cannot be greater than that of Mechanism 1 without violating ex ante IR: in expectation, *all* of the surplus ends up in the hands of the center. That is likely to lead to more instances of "participation regret" in practice than would a mechanism that distributes the surplus more equitably.

Mechanism 2 (surplus sharing)**.** *A mechanism* (f^*, T)*, where,* $\forall i \in I$ *and* $\theta \in \Theta$*:*

$$T_i(\theta) = v_{-i}(\theta) - \frac{n}{n+1}\, \mathbb{E}_{\tilde{\theta}_i}[v(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] \quad (4)$$

Theorem 4. *For all distributions over types, Mechanism 2 is truthful and efficient in dominant strategies, ex ante* individually rational, and ex ante no-deficit. The a priori expected payoff to each agent and the center is identical in the truthful equilibrium.*

3.3 A parameterized variant for trade settings

Consider an asymmetric setting, for instance where there is an initial-holder of a resource that, initially, is of uncertain value (imagine, e.g., start-up companies and a segment of wireless spectrum or plot of land containing rare essential natural resources). To efficiently reallocate the resource it is essential that the initial-holder and the center participate. Here we design a parameterized mechanism with that fact in mind, providing incentives for other agents as well, and allowing for a balance to be customized. The mechanism has parameter ρ which defines the expected percentage share of the social surplus that the initial-holder and the center each get.[8]

Mechanism 3 (risk minimizing)**.** *With parameter* $\rho \in \Re$*, for settings with an agent* h *who is the initial-holder of items for reallocation. A mechanism* (f^*, T)*, where,* $\forall \theta \in \Theta$*:*

$$T_h(\theta) = v_{-h}(\theta) - (1-\rho) \cdot \mathbb{E}_{\tilde{\theta}_h}[v(\tilde{\theta}_h, \theta_{-h}) \,|\, \theta_{-h}], \text{ and} \quad (5)$$

$$T_{i \neq h}(\theta) = v_{-i}(\theta) - \frac{n-2+2\rho}{n-1} \cdot \mathbb{E}_{\tilde{\theta}_i}[v(\tilde{\theta}_i, \theta_{-i}) \,|\, \theta_{-i}] \quad (6)$$

Theorem 5. *For all distributions over types,*[9] *for all* $\rho \in [0, 0.5]$*, Mechanism 3 is truthful and efficient in dominant strategies, ex ante* individually rational, and ex ante no-deficit.*

As we consider the Mechanism 3 schema, a natural optimization problem arises. Ideally neither the center nor any agent ends up worse off for having participated. We know

[7] Proofs for all results subsequent to Theorem 1 are omitted here due to lack of space.

[8] Mechanism 3 has Mechanism 2 as a special case; the two coincide when ρ is set to $\frac{1}{n+1}$ in Mechanism 3.

[9] Note that the theorem holds in all settings; h can be identified arbitrarily in non-trade environments. The mechanism was designed to be *particularly* efficacious for asymmetric settings like trade.

this to be impossible to achieve *universally*; however, we can tailor our mechanism to try to achieve it as often as possible. Specifically, given a problem domain defined by distributions over agent types, we can optimize the choice of parameter ρ. By Theorem 5, whatever our choice of $\rho \in [0, 0.5]$, we will achieve ex ante* IR and ex ante no-deficit. Let $\mathfrak{u}_i^\rho(\theta)$ denote the dominant strategy (truthful) equilibrium payoff to agent i when type profile θ is reported, and let $\mathfrak{u}_c^\rho(\theta)$ be the revenue. We can set:

$$\rho \in \arg\min_{\rho' \in [0,0.5]} \max_i \{\Pr_{\tilde{\theta}}(\mathfrak{u}_i^{\rho'}(\tilde{\theta}) < 0),\ \Pr_{\tilde{\theta}}(\mathfrak{u}_c^{\rho'}(\tilde{\theta}) < 0)\} \quad (7)$$

Another optimization criterion, which we adopt in the next section in forming Figure 2, maximizes the loss-aversion coefficient that still leads to participation for every agent.

4 Evaluation

Our primary goal in this paper was to design efficient mechanisms geared towards maximizing participation. In the context of agents that are not risk-neutral, but rather exhibit *loss-aversion*, this goal is tightly tied to minimizing the probability of utility loss. We consider the following simple and widely-adopted ([Fehr and Goette, 2007] is a recent example) model of loss-averse utility. Each agent i has loss-aversion coefficient $\lambda_i \geq 1$ and, when i obtains value plus payment equal to $x \in \Re$, his experienced utility is:

$$u_i(x) = \begin{cases} x & \text{if } x \geq 0 \\ \lambda_i \cdot x & \text{if } x < 0 \end{cases} \quad (8)$$

$\lambda_i = 1$ is the risk-neutral case, and as λ_i grows agents will be less and less likely to participate in a mechanism that brings significant probability of loss. We now evaluate the mechanisms proposed in the last section, first in terms of raw loss probabilities, then in light of this loss-averse behavioral model, for population size ranging from 2 to 20. We specifically address the following questions:

1. How do the mechanisms compare in terms of probability of utility loss for agents and probability of deficit for the center? (Figure 1)
2. What is the maximum loss-aversion coefficient, held by any agent or the center, for which each mechanism will continue to be ex ante* IR? I.e., given that agent utilities are as in Eq. (8), what is the loss-aversion point at which some agent or the center would ex ante opt not to participate, thus precluding efficiency of the mechanism? (Figure 2)

The results of these queries will depend on the distribution over agent types that one assumes. We considered uniform and normal distributions, for both symmetric settings (where all agents' surplus values are drawn the same way) and trade settings (where one agent only stands to lose his value). The distinction between uniform and normal distributions was minor, and for space reasons we present only the uniform values single-item trade case here. The quantities we wish to consider are complex enough to make analytical evaluation impractical, so we computed estimates with a Monte Carlo sampling method.[10]

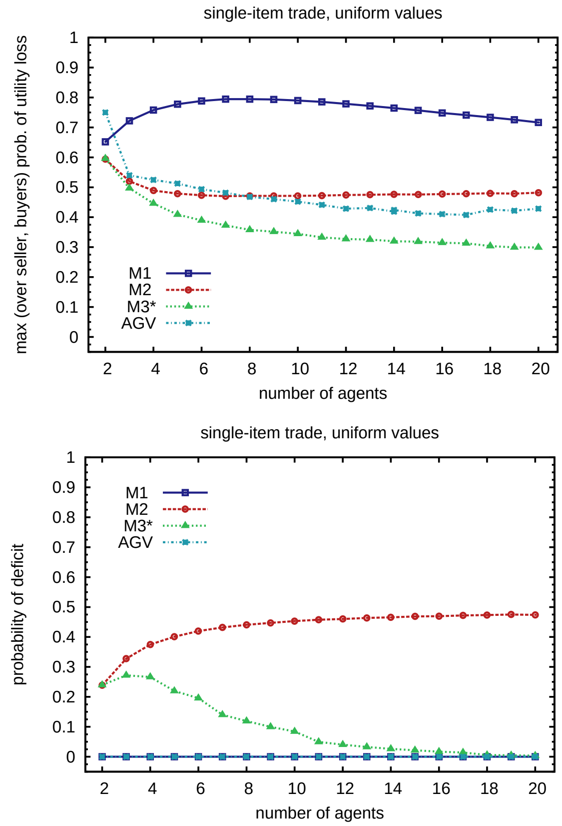

Figure 1: Comparison of mechanisms by loss-probability in a uniform values trade setting. Mechanisms 1 and AGV never yield a deficit, while Mechanism 3 is superior in avoiding agent utility loss. "M3*" represents the instance of the Mechanism 3 class satisfying Eq. (7).

For query (1.) we include results for AGV as a reference point, although it's a poor reference point (it has an "unfair advantage") because it fails to generally satisfy ex ante IR (it is ex ante IR for this setting for population size 3 or greater) and is efficient only in the much weaker Bayes-Nash equilibrium, and only for risk-neutral agents. AGV is not a valid comparison point for query (2.) at all because with loss-averse agents it is not even Bayesian incentive compatible. It is the strategyproofness of our proposed mechanisms that makes them robust to different utility models, including loss-aversion. Thus the most meaningful comparison is between Mechanisms 1, 2, and 3 introduced in this paper.

With respect to query (1.), we see in Figure 1 that the optimized Mechanism 3 (with ρ between 0.32 and 0.4 depending on population size, here) yields the least loss probability for agents—even less than that of AGV; it also achieves a low probability of deficit that converges to 0 as the number of agents grows.

[10]For Figure 2 data-points were computed via a binary search over loss-aversion coefficients to get a course estimate of the point λ where utility transitions from positive to negative, coupled with inspection of expected utility estimates in the space around λ to more precisely identify the 0 point.

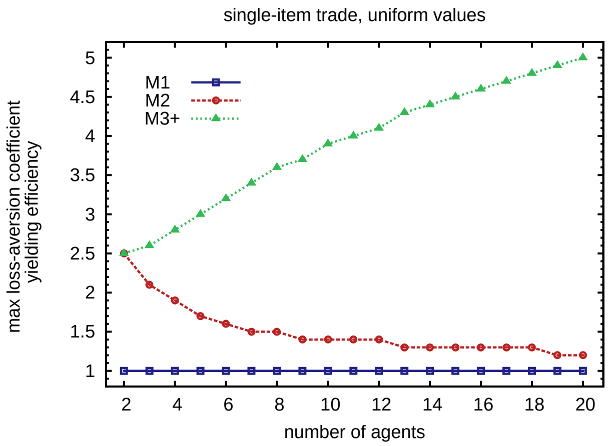

Figure 2: Maximum loss-aversion coefficient for which each mechanism remains efficient, as a function of the number of agents. For a trade setting with uniform values. "M3+" is defined to maximize feasible loss-aversion coefficient.

With respect to (2.), the results are stark. We see in Figure 2 that the optimum of the Mechanism 3 class is very effective and far superior to the other mechanisms. Since in Mechanism 1 expected utility is 0 for all agents, any positive loss-aversion will preclude participation; Mechanism 2 gets more risky for the initial-holder of the good (and less risky for others) as population size grows, even though expected utility increases for all. Mechanism 3 alone (with $\rho \approx 0.35$, here) gets less risky with population size for all agents. Even in a setting with only 5 agents, if all agents and the center weigh the utility of lost value 3 times as much as gained value ($\lambda = 3$), all will choose to participate under Mechanism 3 but will not come close to participation in the other mechanisms. Tversky and Kahneman [1992] experimentally find loss-aversion coefficient λ about 2.25, which Mechanism 3 clears for all population sizes.

5 Discussion

In the introduction we mentioned that a significant amount of work in mechanism design has, in the face of the Myerson-Satterthwaite impossibility theorem, either made strong distribution assumptions or sacrificed efficiency in order to achieve interim individual rationality. There is often good reason to prioritize IR in this way: a clever mechanism isn't any use if you can't get agents to show up for it. Particularly in resource "reallocation" problems, where an agent is the initial holder of a good to potentially be traded, it is often natural that the good holder would know his value and simply not participate if he expects to lose out from doing so.

Of course when participation can be forced, the results of this paper are useful for minimizing ex post regret and grievance among coerced parties. But more importantly, there are scenarios where expecting to benefit *before* learning your value is sufficient for voluntary participation. One example has to do with new goods, where even the initial owner doesn't have private information about his value at first, and faces a narrow window of opportunity for trade. Another natural setting is that of agreements that are extended over time, as in the collaboration example we began the paper with. In such cases a social planner with decision-making authority who prioritizes efficiency but is unwilling to subsidize decisions (though he may be willing to incur some small risk) can propose a "take it or leave it" offer for agents to participate in a decision process—a contractual obligation decided upon before learning types. Agents will participate if and only if expected outcomes fall within their risk tolerances. This paper is distinguished from previous work by addressing such "risky participation" settings and the implications of different risk attitudes therein. We demonstrated that solutions tailored in the style of Mechanism 3—coupled with the analysis quantifying and limiting the probability of loss—will elicit participation even from agents that are significantly loss-averse.

References

[Arrow, 1979] Kenneth J. Arrow. The property rights doctrine and demand revelation under incomplete information. In M. Boskin, editor, *Economics and Human Welfare*. Academic Press, 1979.

[Clarke, 1971] Edward Clarke. Multipart pricing of public goods. *Public Choice*, 8:19–33, 1971.

[Cremér and McLean, 1988] Jacques Cremér and Richard McLean. Full extraction of the surplus in bayesian and dominant strategy auctions. *Econometrica*, 56:1247–1257, 1988.

[D'Aspermont and Gerard-Varet, 1979] C. D'Aspermont and L.A. Gerard-Varet. Incentives and incomplete information. *Journal of Public Economics*, 11:25–45, 1979.

[D'Aspremont *et al.*, 1993] Claude D'Aspremont, Jacques Crémer, and L. A. Gerard-Varet. Correlation, independence, and bayesian implementation. IDEI Working Paper, n. 26, 1993.

[DellaVigna, 2009] Stefano DellaVigna. Psychology and economics: Evidence from the field. *Journal of Economic Literature*, 47(2):315–372, June 2009.

[Fehr and Goette, 2007] Ernst Fehr and Lorenz Goette. Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review*, 97(1):298–317, March 2007.

[Gresik and Satterthwaite, 1989] T. Gresik and M. Satterthwaite. 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*, 48:304–332, 1989.

[Groves, 1973] Theodore Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[Holmstrom, 1979] Bengt Holmstrom. Groves' scheme on restricted domains. *Econometrica*, 47(5):1137–1144, 1979.

[Jackson, 2000] Matthew O. Jackson. Mechanism theory. In *The Encyclopedia of Life Support Systems*. EOLSS Publishers, 2000.

[Maskin and Riley, 1984] Eric S. Maskin and John G. Riley. Optimal auctions with risk averse buyers. *Econometrica*, 52(6):1473–1518, November 1984.

[McAfee and Reny, 1992] R. Preston McAfee and Philip J. Reny. Correlated information and mechanism design. *Econometrica*, 60:395–421, 1992.

[Morris, 1995] Stephen Morris. The common prior assumption in economic theory. *Economics and Philosophy*, 11:227–253, 1995.

[Myerson and Satterthwaite, 1983] Roger Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 28:265–281, 1983.

[Rustichini *et al.*, 1994] Aldo Rustichini, Mark A. Satterthwaite, and Steven R. Williams. Convergence to efficiency in a simple market with incomplete information. *Econometrica*, 62:1041–1064, 1994.

[Sundararajan and Yan, 2010] Mukund Sundararajan and Qiqi Yan. Robust mechanisms for risk-averse sellers. In *Proceedings of the 11th ACM conference on Electronic commerce*, pages 139–148, 2010.

[Tatur, 2005] Tymon Tatur. On the trade off between deficit and inefficiency and the double auction with a fixed transaction fee. *Econometrica*, 73:517–570, 2005.

[Tversky and Kahneman, 1992] Amos Tversky and Daniel Kahneman. Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, 5(4):297–323, 1992.

[Vickrey, 1961] William Vickrey. Counterspeculations, auctions, and competitive sealed tenders. *Journal of Finance*, 16:8–37, 1961.

[Yoon, 2008] Kiho Yoon. The participatory Vickrey–Clarke–Groves mechanism. *Journal of Mathematical Economics*, 44:324–336, 2008.