Project N^o: **FP7-284731**

Project Acronym: **UaESMC**

Project Title: **Usable and Efficient Secure Multiparty Computation**

Instrument: **Specific Targeted Research Project**

Scheme: **Information & Communication Technologies**

Future and Emerging Technologies (FET-Open)

Deliverable D3.2
Strongly Truthful and Composable Mechanism Design

Due date of deliverable: 31st January 2013

Actual submission date: 31st January 2013

Start date of the project: **1st February 2012** Duration: **36 months**

Organisation name of lead contractor for this deliverable: **UoA**

Specific Targeted Research Project supported by the 7th Framework Programme of the EC		
Dissemination level		
PU	Public	✓
PP	Restricted to other programme participants (including Commission Services)	
RE	Restricted to a group specified by the consortium (including Commission Services)	
CO	Confidential, only for members of the consortium (including Commission Services)	

Executive Summary:
Strongly Truthful and Composable Mechanism Design

This document summarizes deliverable D3.2 of project FP7-284731 (UaESMC), a Specific Targeted Research Project supported by the 7th Framework Programme of the EC within the FET-Open (Future and Emerging Technologies) scheme. Full information on this project, including the contents of this deliverable, is available online at `http://www.usable-security.eu`.

In this report we are mostly presenting results from our recent paper [6]. We are presenting a new, extended utilities model for mechanism design settings which can incorporate *externalities*, such as malicious and spiteful behavior of the participating agents. Based on this, a new notion of *strong* truthfulness is presented and analyzed, which is based on the principle of "punishing" players that lie. Due to this, strongly truthful mechanisms can serve as subcomponents in bigger mechanism protocols in order to "boost truthfulness" in settings with externalities and achieve a kind of externalities-resistant performance. The related solution concept equilibria are discussed. This decomposability scheme is rather general and powerful, and we show how it can be also adapted to the case of the well known mechanism design problem of scheduling tasks to machines for minimizing the makespan.

List of Authors

Yiannis Giannakopoulos (UoA)

Contents

Chapter 1

Foundations

In this chapter we are going to present some fundamental notions from the area of Mechanism Design (MD), upon which we are going to build our exposition and our results in the subsequent chapters. We deliberately choose to make this presentation based on a simple, single-item auction paradigm, in order to demonstrate more clearly both the intuition and the essence of these notions and not get lost into the technicalities of the general models which, after all, we are not going to need for the purposes of this deliverable. The reader of course can find many very good references including more formal and general introductions to the fascinating area of (Algorithmic) Mechanism Design (see e.g. [9, 10]).

1.1 Basic Notions from Game Theory and Mechanism Design

Assume the following traditional mechanism design setting, in particular a single-item auction scenario: We have n players (also called agents), each of whom is willing to pay t_i, $i = 1, 2, \ldots, m$, in order to get the item. These are called the agents' *types*, and $\boldsymbol{t} = (t_1, t_2, \ldots, t_n)$ is the *type profile*. Lets assume that all these belong to some domain T, where in our particular auction setting it's a natural assumption to consider $T = \mathbb{R}_+$. This is *private* information of the players, who report it to the auctioneer in the form of *valuations* (also called *bids* in the context of auctions) v_i, $i = 1, 2, \ldots, m$. The reason we discriminate between types and valuations is that the players, as we will see shortly, may have a reason to lie about their true types and misreport some $v_i \neq t_i$. Given the input by the players, i.e. the bid profile $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$, the auctioneer need to decide who gets the item and how much is she going to pay for that. More formally, a *mechanism* $\mathcal{M} = (\boldsymbol{a}, p)$ comprises of an *allocation* vector $\boldsymbol{a} = \boldsymbol{a}(\boldsymbol{v}) = (a_1(\boldsymbol{v}), a_2(\boldsymbol{v}), \ldots, a_n(\boldsymbol{v})) \in [0, 1]^m$ and a *payment* vector $\boldsymbol{p} = \boldsymbol{p}(\boldsymbol{v}) = (p_1(\boldsymbol{v}), p_2(\boldsymbol{v}), \ldots, p_n(\boldsymbol{v})) \in \mathbb{R}_+^n$. If our mechanism is a deterministic one, $a_i(\boldsymbol{v})$ is simply an indicatior 0–1 variable denoting whether player i wins the item or not, while if we allow for randomized mechanisms, $a_i(\boldsymbol{v})$ denotes the probability of player i winning the item. In this latter case, we must make sure that $\sum_{i=1}^{n} a_i(\boldsymbol{v}) \leq 1$ for all $\boldsymbol{v} \in T^n$. Also, agent i will have to submit a payment of $p_i(\boldsymbol{v})$ to the mechanism. We define player's i *utility* to be his total "happiness" after taking part to the auction minus the payment he has submitted, formally:

$$u_i(\boldsymbol{v}|t_i) = a_i(\boldsymbol{v}) \cdot t_i - p_i(\boldsymbol{v}). \tag{1.1}$$

Notice the notation $u_i(\boldsymbol{v}|t_i)$ and the different usage of valuations and types in expression (1.1). In case of *truth-telling*, i.e. honest reporting of $v_i = t_i$, we will simplify notation to

$$u_i(\boldsymbol{v}) = a_i(\boldsymbol{v}) \cdot v_i - p_i(\boldsymbol{v}). \tag{1.2}$$

We call these *quasilinear* utilities, due to the special form of these expressions, and in particular the linear-form connection between a player's utility and *only her own* type t_i.

Surprisingly enough, if we look a little closer we can see that we have already formed a *game* (see e.g. Chapter 2 of Deliverable 3.1): the players' strategies are exactly their valuations and their goal is to, each one of them, being completely rational, to *selfishly* maximize her own utility. So, we can use standard solution concepts and in particular *equilibria* in order to talk about possible stable states of our auctions. For example, the most fundamental notion that underlines the entire area of Mechanism Design, is that of *truthfulness* (also called sometimes incentive compatibility or strategyproofness). Intuitively, we will say that a mechanism is truthful if it makes sure that no agent has an incentive to lie about her true type:

Definition 1.1.1 (Truthfulness)**.** *A mechanism $\mathcal{M} = (\boldsymbol{a}, \boldsymbol{p})$ is called truthful if truth-telling is a* dominant-strategy *equilibrium of the underlying game, i.e. $u_i(v_i, \boldsymbol{v}_{-i}|v_i) \geq u_i(\tilde{v}_i, \boldsymbol{v}_{-i}|v_i)$ for every player i, all possible valuation profiles $\boldsymbol{v} \in T^n$ and all possible "misreports" $\tilde{v}_i \in T$.*

In case of randomized mechanisms, the above definitions are being naturally extended by taking expectations of the utilities (*truthful in expectation* mechanisms).

Being implemented in dominant strategies, truthfulness is a very stable and desirable property, which we want all our auction mechanisms to satisfy. It allows us to "extract the truth" from the participating parties and thus be able to accurately design the protocols for the goals we want to achieve. A celebrated result by Myerson gives us a powerful and simple characterization of truthful mechanisms and also helps us completely determine a mechanism simply by giving just its allocation function $\boldsymbol{a}$:

Theorem 1.1.2 (Myerson [8])**.** *A mechanism $\mathcal{M} = (\boldsymbol{a}, \boldsymbol{p})$ is truthful if and only if*

1. *Its allocation functions are monotone nondecreasing, in the sense that*

$$v_i \leq v_i' \quad \Longrightarrow \quad a_i(v_i, \boldsymbol{v}_{-i}) \leq a_i(v_i', \boldsymbol{v}_{-i})$$

for every player i, all valuation profiles $\boldsymbol{v}_{-i}$ and all valuations v_i, v_i'.

2. *The payment functions are given by:*

$$p_i(\boldsymbol{v}) = a_i(\boldsymbol{v})v_i - \int_0^{v_i} a_i(x, \boldsymbol{v}_{-i})\, dx. \tag{1.3}$$

Based on this, one can show an other elegant and concise, analytic characterization of truthful mechanisms:

Theorem 1.1.3. *A mechanism is truthful if and only if the players' utilities that it induces are* convex *functions (over T^n).*

1.2 Fundamental MD Problems

In this section we very briefly present two of the most fundamental mechanism design domains in Algorithmic Game Theory, as well as the associated optimization objectives. These are the predominant motivating examples that move the entire field forward and also serve as the basis for our exposition in this deliverable both technically and in spirit.

1.2.1 Auctions: Welfare and Revenue

The fundamental single-item auction introduced in Section 1.1 can be generalized to the following m items *additive* valuations auction setting where now the allocation $\boldsymbol{a}$ of a mechanism $\mathcal{M} = (\boldsymbol{a}, \boldsymbol{p})$ is an $n \times m$ matrix $\boldsymbol{a} = \{a_{ij}\} \subseteq [0,1]^{n\times m}$ where a_{ij} represents the probability of agent i getting item j, $= 1, 2, \dots, n$, $j = 1, 2, \dots, m$. The inputs to the mechanism are bid profiles $\boldsymbol{b} \in T^{n\times m}$, $T = \mathbb{R}_+$, where b_{ij} is player's i valuations for item j. Of course, we must make sure that for every possible input $\boldsymbol{b}$ the selected outcome $\boldsymbol{a}(\boldsymbol{b})$ must not assign any item to more than one agent, i.e.

$$\sum_{i=1}^{n} a_{ij}(\boldsymbol{b}) \leq 1 \quad \text{for all items } \; j = 1, 2, \dots, m. \tag{1.4}$$

When we design auction mechanisms we are usually interested in maximizing either the combined "happiness" of our society, meaning the sum of the valuations of the players that receive items, or the auctioneer's profit, i.e. the sum of the payments he collects from the participating agents. So, we define the social *welfare* of mechanism $\mathcal{M}$ on input $\boldsymbol{b}$ to be

$$W^{\mathcal{M}}(\boldsymbol{b}) = \sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}(\boldsymbol{b}) b_{ij}$$

and its *revenue*

$$R^{\mathcal{M}}(\boldsymbol{b}) = \sum_{i=1}^{n} p_i(\boldsymbol{b}).$$

The most well-known auction is without doubt the VCG auction, named after the work of Vickrey [12], Clarke [4] and Groves [7]. In the simple single-item setting, this mechanism reduces to the simple Vickrey second-price auction which gives the item to the highest bidding agent but collects as payment the second-highest bid. In that way, it ensures truthfulness by not giving an incentive to the wining agent to misreport a lower bid, since her payment is independent of her own bid and depends only on the bids of the other players. And, above all, this mechanism maximizes social welfare "by definition".

Formally, generalizing these ideas to the setting of m items, the VCG auction is the mechanism with allocation rule

$$a(\boldsymbol{b}) = \operatorname*{argmax}_{\alpha \in A} \sum_{i=1}^{n} v_i(\alpha(\boldsymbol{b})) = \operatorname*{argmax}_{\alpha \in A} \sum_{i=1}^{n} \sum_{j=1}^{m} \alpha_{ij} b_{ij} \quad \text{for all } \boldsymbol{b} \in T^{n \times m}, \tag{1.5}$$

and payments

$$p_i(\boldsymbol{b}) = \max_{\alpha \in A} \sum_{j \neq i} v_j(\alpha(\boldsymbol{b})) - \sum_{j \neq i} v_j(a(\boldsymbol{b})) \quad \text{for all } \boldsymbol{b} \in T^{n \times m}. \tag{1.6}$$

The underlying idea is that, first of all the allocation rule (1.5) ensures that the social welfare is the optimal one, and then the payments (1.6) are such that they "internalize the externalities" of every player i, that is, intuitively, we charge player i the "harm" that her participation in the auction causes to the rest of the society.

1.2.2 Scheduling: Minimizing Makespan

The Scheduling domain is essentially an additive multi-item auction setting (see the previous Section 1.2.1) with the modification that players are not trying to maximize their utilities, but to *minimize* them: they are "cost" functions. This of course results in a different model, and the exact nature of the relation between the two domains is not completely clear at the moment. The mechanism design version of the problem has been first studied by Nisan and Ronen in their seminal paper [9] and has arguably been the most influential paper in the area of algorithmic mechanism design. So, in Scheduling settings we have n machines (players) and m tasks (items). Each machine reports to the mechanism designer the time she would need to process every item, in the form of a time matrix (type profile) $\boldsymbol{t} = \{t_{ij}\} \subseteq T^{n \times m}$, $T = \mathbb{R}_+$, where t_{ij} is the processing time of machine i for task j. A feasible allocation is an assignment of tasks to players, given (exactly as in the case of additive auctions) by an allocation matrix $\boldsymbol{a} = \{a_{ij}\} \in \{0,1\}^{n \times m}$, $a_{ij} = 1$ if and only if machine i executes task j. Machine's i total valuation is the sum of the processing times for each individual task assigned (additive valuations) $\sum_{j=1}^{m} a_{ij} t_{ij}$. Each machine's resulting cost is $c_i(\boldsymbol{t}) = \sum_{j=1}^{m} a_{ij}(\boldsymbol{t}) t_{ij} - p_i(\boldsymbol{t})$, where $p_i(\boldsymbol{t})$ represents the payments with which we compensate machine i in order to "motivate" her to take part in the execution.

Especially with the emergence of the Internet as the predominant computing paradigm, it is natural to assume that these "machines" would act selfishly, just caring about minimizing their own costs $c_i(\boldsymbol{t})$, and possibly misreport their true processing times towards this ends. So, a game theoretic approach to the classical task allocation problem under the *added truthfulness* constraint is both interesting and necessary. The standard objective is to design truthful mechanisms that minimize the *makespan*

$$\text{Makespan}(\boldsymbol{t}) = \max_i \sum_{j=1}^{m} a_{ij}(\boldsymbol{t}) t_{ij},$$

that is, the time it would take the slowest machine to finish the processing. Again, we will also require that no task is unprocessed, thus (1.4) becomes

$$\sum_{i=1}^{n} a_{ij}(\boldsymbol{t}) = 1 \quad \text{for all tasks } \ j \ \text{ and time matrices } \ \boldsymbol{t} \in T^{n\times m}. \tag{1.7}$$

This is known as the *Scheduling problem in parallel unrelated machines.* When we consider the *fractional allocations* variant of the scheduling problem, we allow $\{a_{ij}\} \in [0,1]^{n\times m}$, while still demanding condition (1.7). Fractional allocations are essentially randomized mechanisms for the non-fractional version of the problem.

Chapter 2

Externalities and Challenges

In this chapter we describe the notion of *externalities* in mechanism design settings and present the game theoretic and algorithmic challenges that they introduce. This discussion is inevitably intertwined with challenging the fundamental notion of traditional *truthfulness* itself, and thus this chapter forms essentially the motivation for our main results of this deliverable, presented in Chapter 3.

2.1 Externalities in Mechanism Design

So, what are externalities? A fundamental assumption throughout Game Theory and Mechanism Design is that all participating players (agents) are fully rational and act *selfishly*: they have their own well-defined utility function (see e.g. (1.2)) and they only care about optimizing this function. This can be maximizing "satisfaction" when they want to buy an item in an auction, or minimizing their cost when they are machines that get to execute tasks (scheduling problem, see section 1.2.2). Usually, this utility function optimization is considered myopic, in the sense that players do not care about variations on other players' achieved utilities as far as these variations do not effect their own utility levels. For example in a standard single-item auction setting, if some player does not get the item (thus achieving a zero utility) she is indifferent (i.e. her utility does not change) towards how much the winning bidder is going to pay for obtaining the item and, even more, she doesn't care about this bidder's identity.

But is this really a "natural" or "expected" behavior when we think of everyday social interactions? Experiments show that bidders can overbid, possibly risking of ending up with negative utilities, just for the "joy of winning" in case they get the item. Or, on the other end, if they do not manage to win the item, overbidding will drive prices up in order to "harm" the other winning agent(s), an arguably "spiteful" behavior.

The above demonstrate examples where participants in a mechanism behave seemingly irrationally, their happiness not only being a function of just their own "core" utility but also being affected in complex ways by the other players' utilities. We will call such effects on the modeling of our agents' happiness *externalities*, to emphasize the "third-party" nature of the interaction. Of course, externalities are not only *negative* like the examples we gave before. They can be *positive*, "altruistic" in nature, e.g. a loving couple taking part at the same auction: one partner may be happy to lose the item if it is for the other one to get it.

Externalities have been heavily studied in economics, not just Game Theory, and the literature is extensive and diverse. For our purposes, we will approach externalities in the context of the informal definition we gave in the previous paragraphs.

2.2 Impact on Truthfulness

Under the scope of externalities, let us revisit the canonical example of a single-item auction. We know that the second-price paradigm, in particular the Vickrey auction, which gives the item to the highest bidding agent but collects as payment the second-highest bid, is optimal for social welfare while also being truthful:

no agent has an incentive to lie about her true valuation (bid) *no matter what* the other players report (i.e., in the language of Game Theory, truth-telling is a dominant strategy equilibrium). A magnificent and elegant result, that allows us not only to maximize the "collective happiness" of the participants but also ensure the integrity of the extracted information, here the agents' bids. Furthermore, this is backed-up by a very powerful notion of "stability" for our solution, that of dominant strategies implementation.

But if we consider spiteful behavior, this no more holds: the second highest bidder, who is going to lose the item anyway, can harm the winning agent by declaring an (untruthful) higher bid which immediately ("externally") affects the payment of the other players, and in particular increases the payment for the winning player. Is this desired? Is it expected? Can we model, study and prevent this phenomena? One simple suggestion would be to "encode" such possible externalities-behavior into each player's type (some kind of "generalized" valuation) and then run the powerful VCG mechanisms (see Section 1.2.1) on the new extended type-profile space of all agents to get out of the box again a socially efficient and dominant strategy truthful mechanism. There is though a fundamental problem in this approach that conflicts all existing tools in the area of Mechanism Desing: the players' utilities now also depend on *other* agents' payments[1]. To be more precise, utilites are no more *quasi-linear* functions with respect to payments.

2.3 Challenges

Given the above discussion, now the challenges are:

- How to incorporate properly these externalities into the existing standard utility maximization framework of Mechanism Design? We need a new model for the utility functions, taking into consideration both the "core" utilities of each players, but also the "external" ones.
- Is the standard notion of truthfulness, and in particular that of implementation in dominant strategies, computationally and descriptively appropriate to model this new complex framework of interactions? Can we propose a new notion of "empowered" truthfulness?
- Utilizing stronger notions of truthfulness, can we design mechanisms that manage to somehow resist against externalities which threaten the integrity of traditional truthfulness, the building block of the entire area of Mechanism Design?
- Is there a general Mechanism Design paradigm that provides construction of such externality-resistant mechanisms?

These are the main challenges and questions that motivated us throughout the current deliverable. In the following Chapters 3 and 4 we try to make a step towards this direction.

[1]This is in fact a more "appropriate" definition, economics-wise, of externalities themselves: utilities are *externaly* affected by payments to third parties, upon which we have no direct control.

Chapter 3

Strong Truthfulness and Externality Resistance

Under the scope of the introduction of the notion of externalities in Chapter 2 we now have to start dealing with the challenges discussed in Section 2.3. First of all, under these new extremely complex underlying interactions among players, we need to have a solid building-block, stronger than the traditional notion of truthfulness which, as we already have seen even on the simplest example of single-item auctions, can make the powerful VCG second-price auction fail dramatically. The intuition is that we would like to introduce a more strict notion of truthfulness, where not only players have no reason to lie but they are going to be punished for misreports. We want higher deviations from the true valuations to result in higher reductions to the resulting utilities for the players who deviate:

3.1 The Notion of Strong Truthfulness

Definition 3.1.1 (Strong truthfulness)**.** *A mechanism is called c-strongly truthful, for some $c \in \mathbb{R}$, if it induces utilities u_i such that*

$$u_i(v_i, \boldsymbol{v}_{-i}) - u_i(\tilde{v}_i, \boldsymbol{v}_{-i}) \geq \frac{1}{2} c \left| \tilde{v}_i - v_i \right|, \tag{3.1}$$

for every player i, valuations v_i, $\tilde{v}_i$ and all valuation profiles of the other players $\boldsymbol{v}_{-i}$.

This notion of strong truthfulness is a generalization of the "standard" notion of truthfulness, simply by setting $c = 0$.

Under these definitions we can now prove an elegant analytic characterization of strong truthfulness, in the spirit of Theorem 1.1.3:

Theorem 3.1.2. *A mechanism is c-strongly truthful if and only if the utility functions it induces for every player are all c-strongly convex functions.*

3.1.1 Strongly truthful auctions

Let us give an example of a strongly truthful auction in the simplest case of a single-buyer, single-item auction. Assume a single player with a valuation for the item drawn from some bounded real interval $[L, H]$. We define the following mechanism, which we will call *linear*:

Definition 3.1.3 (Linear mechanism)**.** *The linear mechanism, denoted by* LM*, for the single buyer, single item auction setting has allocation*

$$a(v) = \frac{v - L}{H - L},$$

where the buyer's valuation v for the item ranges over a fixed interval $[L, H]$.

It turns out that this elegant mechanism is in fact the *strongest* possible mechanism one could hope for in this setting:

Theorem 3.1.4. *The linear mechanism is $\frac{1}{H-L}$-strongly truthful for the single buyer, single item auctions setting and this is optimal*[1] *among all mechanisms in this setting.*

Notice that Theorem 3.1.4 holds only in the case where the valuations domain is a (real) *interval* $T = [H, L]$. In case we want to deal with unbounded domains, e.g. $T = \mathbb{R}_+$, we need to define a more flexible notion of *relative* strong truthfulness (see [6, Section 2.2])

3.2 External Utilities Model

In order to model externalities, both positive (altruism) and negative (spite), we are extending our agents' types to include not only their valuations but also some parameters γ that try to quantify the interest/external-effect each agent has upon others. Formally, for every agent $i = 1, 2, \ldots, n$ we redefine its type t_i to be $t_i = (v_i, \boldsymbol{\gamma}_i)$ where $\boldsymbol{\gamma}_i = (\gamma_{i1}, \gamma_{i2}, \ldots, \gamma_{in})$ with γ_{ij}, $j = 1, 2, \ldots, n$ being the *externality parameter* of player i with respect to player j. In fact, parameter γ_{ii} is not needed and we can safely ignore it, setting $\gamma_{ii} = 0$ for all players i, but we still keep it into the formation of γ_i in order to have a vector-consistent notation. The usage of these parameters will become apparent when we will define the utilities in our new externalities model, but the intuition is that negative values of γ_{ij} correspond to player i demonstrating spiteful behavior towards player j, positive values correspond to altruistic behavior and a value of zero for this externality parameter corresponds to lack of externalities towards player j (i.e. player i demonstrates "standard" game theoretic selfish behavior).

Moving on to defining utilities, let $\mathcal{M} = (\boldsymbol{a}, \boldsymbol{p})$ be a mechanism for our new externalities model. Our players have true types $t_i = (v_i, \boldsymbol{\gamma}_i)$ and they submit to the mechanism *bids* (also called *values*) which are (possibly mis-) reports of the valuation component v_i of their true type t_i. Given such a bid profile $\boldsymbol{b}$, mechanism $\mathcal{M}$ again computes the allocation $\boldsymbol{a}(\boldsymbol{b}) = (a_1(\boldsymbol{b}), a_2(\boldsymbol{b}), \ldots, a_n(\boldsymbol{b}))$, $a_i(\boldsymbol{b}) \in [0, 1]$ being the probability that agent i "receives service"[2] and payment vector $\boldsymbol{p}(\boldsymbol{b}) = (p_1(\boldsymbol{b}), p_2(\boldsymbol{b}), \ldots, \boldsymbol{p}_n(\boldsymbol{b}))$, where $p_i(\boldsymbol{b})$ is the payment extracted *from* player i.

In this setting, we define player's i *base utility* under mechanism $\mathcal{M}$, given (true) type $t_i = (v_i, \boldsymbol{\gamma}_i)$ and reported bid vector $\boldsymbol{b}$ to be the "standard"[3] utility

$$u_i(\boldsymbol{b}|t_i) = u_i(\boldsymbol{b}|v_i) = a_i(\boldsymbol{b}) \cdot v_i - p_i(\boldsymbol{b})$$

and then we define her *externality-modified utility* given also the other players' (true) type profiles $\boldsymbol{t}_{-i}$ to be

$$\hat{u}_i(\boldsymbol{b}|\boldsymbol{t}) = \hat{u}_i(\boldsymbol{b}|\boldsymbol{v}, t_i) = u_i(\boldsymbol{b}|v_i) + \sum_{j \neq i} \gamma_{ij} u_i(\boldsymbol{b}|t_j). \tag{3.2}$$

From now on we will refer to this externality-modified utility simply as "utility", since this is going to be the utility notion in our new externalities-included model upon which we will also build our new notions of "truthfulness" and "resistant" mechanisms. Some important observations on the above definitions are in place:

- First notice that, as expected, the base utility $u_i(\boldsymbol{b}|\boldsymbol{t}_i)$ only depends on player's i type $\boldsymbol{t}_i$ and not on the other players' types $\boldsymbol{t}_{-i}$ and, in particular, it depends just on the valuation component v_i of $\boldsymbol{t}_i$ (i.e. the externality parameters $\boldsymbol{\gamma}_i = (\gamma_{i\,1}, \gamma_{i\,2}, \ldots, \gamma_{i\,n})$ do not play any part in these base utilites). So, we can also use the slightly lighter notation $u_i(\boldsymbol{b}|v_i)$ to denote the basic utility. Essentially, this is the component of our new utility which corresponds exactly to the standard definition of utilities in the traditional no-externalities setting of Mechanism Design (see Section 1.1).

[1] Formally, for every c-strongly truthful mechanism in this setting it must be that $c \leq \frac{1}{H-L}$.
[2] In the single-item auction paradigm, this means agent i gets the item.
[3] Look introductory Section 1.1.

- The externality-modified utility needs to depend on the entire (true) type profile $\boldsymbol{t}$ and not just i's component t_i. That is, because the "externalities-induced" term of the definition (3.2) comprises of a sum which ranges across all other players. Furthermore, unlike the base utilities, here we don't just need the valuations v_i but all parameters γ_{ij} for all $j \neq i$. However, from the parameters profile vector $\boldsymbol{t}$ we only need the externalities parameters of player i so, again to make the notation more straightforward, we can write $\hat{u}_i(\boldsymbol{b}|\boldsymbol{v}, \boldsymbol{\gamma}_i)$ instead of $\hat{u}_i(\boldsymbol{b}|\boldsymbol{t})$.

Naturally enough, if $b_i = v_i$ we can denote the base utilities $u_i(\boldsymbol{b}|\boldsymbol{t}_i)$ simply by $u_i(\boldsymbol{b})$ and if $\boldsymbol{b} = \boldsymbol{t}$ we can use just $\hat{u}_i(\boldsymbol{b})$ instead of $\hat{u}_i(\boldsymbol{b}|\boldsymbol{t})$ for the externality-modified utilities.

3.3 Externality Resistant Mechanisms

Definition 3.3.1 (Externality-resistant VCG)**.** *The externality resistant-VCG mechanism for our mechanism design setting with externalities, denoted by* rVCG(δ) *and parametrized by some* $\delta \in [0,1]$ *is the following protocol:*

- *Ask all n players to report their bids b_i, $i = 1, 2, \ldots, n$.*
- *With a probability of $\frac{\delta}{n}$ for every player i, $i = 1, 2, \ldots, n$, single-out player i and run the* LM *mechanism (see Definition 3.1.3).*
- *With the complementary probability $1 - \delta$ run the "standard" VCG mechanism.*

Theorem 3.3.2. *Consider a mechanism design setting with externalities where players' valuations are drawn from the unit real interval, i.e. $v_i \in [0,1]$ for all $i = 1, 2, \ldots, m$. Then, for every $\delta, \varepsilon > 0$, if the the externality parameters of our agents satisfy*

$$\max_{i,j} \gamma_{i\,j} < \frac{\varepsilon\delta}{8(1-\delta)^2 n^3},$$

the externality-resistant VCG mechanism (see Definition 3.3.1) induces utilities such that

$$u_i^{\mathrm{rVCG}(\delta)}(\boldsymbol{b}) \geq (1-\delta) u_i^{\mathrm{VCG}}(\boldsymbol{v}) - \varepsilon$$

for every player $i = 1, 2, \ldots, n$ and all undominated *strategy (bid) profile $\boldsymbol{b}$ of the players.*

The intuition behind this result is that, by randomizing over components/mechanisms that are known to work in simpler settings, the rVCG mechanism manages to achieve (base) utilities that are very close to these of the corresponding VCG mechanism in the no-externalities setting (see Section 1.2.1). Of course, this is not implementable in dominant strategies (we have seen that truthfulness under so strong solution concepts are doomed to fail) but under a weaker solution concept, that of undominated strategies (see next Section 3.4).

As an immediate result of this virtual "simulation" of an ideal behavior where only base utilities are taken into consideration, rVCG manages to approximate both optimal social welfare as well as the revenue of the traditional VCG (run on base utilities):

Theorem 3.3.3. *In every outcome of* rVCG(δ) *implemented in undominated strategies,* rVCG(δ)*'s social welfare is within an additive error of $n\eta$ from the optimal social welfare and within an additive error of $2n\eta$ of the revenue achieved by the standard* VCG *mechanism (run on base utilies without externalities), where n is the number of players and η is a parameter with*

$$\eta \leq \frac{4(1-\delta)}{\delta} n^2 \gamma$$

(where $\gamma = \max_{i,j} \gamma_{ij}$).

3.4 Implementation in Undominated Strategies

Let us describe briefly the solution concept we use in our model and under which our notion of externality resistance is realized, namely *implementation in undominated strategies.* Intuitively, we say that a given property P (here, "externality resistance") is implemented in undominated strategies if, for every agent i there exists a set of strategies D_i such that "playing within D_i" is a kind of "dominant strategy" for every player i, i.e. no matter what the other players' strategies are, there is some strategy in D_i that maximizes player's i utility. In addition, obviously P must be satisfied for all possible input strategies in the product space $\prod_{i=1}^{n} D_i$. For a more formal description we refer to Section 1.4 of our paper [6] as well as [1] that utilized this solution concept before. In our specific model the idea is that as long as the agents stay in strategies which are close enough to truth-telling they are "safe". Deviating out of the "balls" will be an absurd choice, a dominated strategy.

Chapter 4

Protocol Composability

In this chapter we explore the bigger image behind our proposed notion and constructions of the previous Chapter 3. What is the grand underlying schema that achieves this resistance towards externalities while still approximating the mechanism designer's objective (e.g. social welfare, see Theorem 3.3.3)?

4.1 "Boosting" Truthfulness

By looking at Theorem 3.3.2 we see that the key property of rVCG (see Definition 3.3.1) is that the following two values are *approximately* equal for all agents:

- The utility they end up with in the model *with* externalities after running rVCG
- The utility they would have ended up with in an ideal *no*-externalities model after running the "traditional" VCG mechanism.

In other words, while all agents still bid as to maximize their new, complex *externality-modified* utilities, they end up with a base utility that is approximately what it would have been if all agents bid so as to maximize their *base* utility. Thus, these externally-resistant mechanisms try to "simulate" the agents' behavior in an externalities-free "Utopia" and, as a result, they manage to approximate the optimal social welfare and revenue.

But, above all, what these mechanisms achieve is to "boost" truthfulness by enforcing incentive-compatibility in this challenging externalities-modified utility model. How is that achieved? The key design feature is that of *composability*: If we look at rVCG (Definition 3.3.1) we will see that it *randomizes* over *strongly-truthful* mechanisms. In particular, it uses the advantage of a strongly-truthful mechanism that punishes agents for misbehaving in order to "forcefully" extract truthful reporting by the agents. It does so, by running with some (small) probability such a punishing protocol on a random agent. With the remaining probability we run a "standard" truthful mechanism that performs optimally with respect to base utilities. In other words, we enrich mechanisms that perform well in the traditional externalities-free model by *composing* them with "small", powerful, strongly-truthful subroutines.

Such a composable design paradigm, where different mechanisms are combined to boost truthfulness has been used before in mechanism design settings, e.g. utilizing differential privacy [11] and more subtly in the scoring rules [3, 2] and responsive lotteries [5]. However, in our current paper we are making a first step towards the systematic study of this scheme and the quantification of the performance of the composable mechanisms. Also, this is the first time when this design is used to achieve externality resistant. Furthermore, our constraction has the advantage that it is readily applicable to multi-dimensional mechanism design settings, such us multi-item auctions and scheduling jobs to machines, which is the subject matter of the next section.

4.2 Extensions and Multi-dimensional Domains

The externality-resistant mechanism rVCG, presented in Section 3.3, applied in a simple, single-dimensional auction setting with only a single item for sale. Can we extend this powerful externality-resistant idea and the composable scheme that we described in Section 4.1 to incorporate more involved *multi-dimensional* settings with many items, or the scheduling problem?

It turns out that our construction is generic enough to do that in a very similar, straightforward way. Consider for example the mechanism design Scheduling problem of minimizing the makespan of unrelated machines (see Section 1.2.2). We show how to take advantage of strongly truthful mechanisms to give an "unexpected" solution to this problem. We will give a mechanism for the problem under the following assumptions:

- The execution times are bounded; in particular we assume that $t_{i,j} \in [L, H]$.
- As in the classical version of the problem, each task must be executed at least once but in our version it may be executed more than once, even by the same machine [1]. When a machine executes the same task many times, we assume that it pays the same cost t_{ij} for *every* execution.
- The solution concept for truthfulness of our mechanism is not dominant strategies but *undominated strategies* (see 3.4).

The mechanism is defined by two parameters $\delta \in [0, 1]$ and r:

- The players declare their values $\tilde{t}_{ij}$, perhaps different than the real values t_{ij}.
- With probability $1 - \delta$, using the declared values assign the tasks optimally to the players[2].
- With the remaining probability δ, for every player i and every task, run "truth extracting" LM mechanism (see Definition 3.1.3), like in the case of the externality resistant VCG (Definition 3.3.1) r times using the declared values $\tilde{t}_{ij}$ from the first step. (In fact, we need only to simulate LM once, pretending that the execution time of every task has been scaled up by a factor of r.)

We will show that

Theorem 4.2.1. *For every $\delta > 0$ and $\epsilon > 0$, we can choose the parameter r so that with probability $1 - \delta$ the mechanism has approximation ratio $1 + \epsilon$ and makes no payments; the result holds as long as the players do not play a dominated strategy. This, for example, is achieved for every*

$$r \geq 8n^2 m \frac{H^2}{L^2} \frac{1}{\delta \cdot \epsilon^2}.$$

Proof. The main idea is that if a machine lies even for one task by more than $\epsilon_0 = \frac{L}{2n}\epsilon$, the expected cost of the lie in the truth extraction part of the mechanism will exceed any possible gain; therefore, the truth-telling strategy dominates any strategy that lies by more than ϵ_0.

We now proceed with the calculations: If a machine lies about one of its tasks by at least an additive term ϵ_0, it will pay an expected cost of at least $r\delta\frac{1}{2}\frac{\epsilon_0^2}{H-L}$. The maximum gain from such a lie is to decrease (with probability $1 - \delta$) its load from mH (the maximum possible makespan) to 0. So the expected gain is at most $(1 - \delta)mH \leq mH$, while the loss is at least $r\delta\frac{1}{2}\frac{\epsilon_0^2}{H-L}$. If we select the parameters so that

$$r\delta\frac{1}{2}\frac{\epsilon_0^2}{H-L} \geq mH, \tag{4.1}$$

[1] The proofs of Nisan and Ronen that give lower bound of 2 and upper bound of n for the approximation ratio can be easily extended to this variant of the scheduling problem. The same holds for the lower bound of truthful-in-expectation mechanisms.

[2] Finding or even approximating the optimal allocation with a factor of 1.5 is an NP-hard problem, but this is not a concern here since we focus on the game-theoretic difficulties of the problem. One could replace this part with an approximation algorithm to obtain a polynomial-time approximation mechanism.

no machine will have an incentive to lie by more than ϵ_0, i.e., $|\tilde{t}_{i,j} - t_{i,j}| \leq \epsilon_0$. But then the makespan computed by the mechanism cannot be more than $m\epsilon_0$ longer the optimal makespan: $\text{Makespan}(\tilde{t}) \leq \text{Makespan}(t) + m\epsilon_0$. We can use the trivial lower bound $\text{Makespan}(t) \geq mL/n$ (or equivalently $n \,\text{Makespan}(t)/(mL) \geq 1$) to bound the makespan of $\tilde{t}$:

$$\begin{aligned}
\text{Makespan}(\tilde{t}) &\leq \text{Makespan}(t) + m\epsilon_0 \\
&\leq \text{Makespan}(t) + m\epsilon_0 \frac{n \,\text{Makespan}(t)}{mL} \\
&= \left(1 + \frac{n\epsilon_0}{L}\right) \text{Makespan}(t) \\
&= (1 + \epsilon) \,\text{Makespan}(t),
\end{aligned}$$

where $\epsilon = \frac{n\epsilon_0}{L}$. We can make the value of ϵ as close to 0 as we want by choosing an appropriately high value for r; Constraint (4.1) shows that $r = \Theta(\delta\epsilon^{-2})$ is enough. Therefore, with probability $1 - \delta$, the makespan of the declared values is $(1 + \epsilon)$-approximate, for every fixed $\epsilon > 1$. □

Bibliography

[1] M. Babaioff, R. Lavi, and E. Pavlov. Single-value combinatorial auctions and implementation in undominated strategies. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 1054–1063. ACM, 2006.

[2] J. Eric Bickel. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, 4(2):49–65, June 2007.

[3] G.W. Brier. Verification of forecasts expressed in terms of probability. *Monthly weather review*, 78(1):1–3, 1950.

[4] E.H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

[5] U. Feige and M. Tennenholtz. Responsive lotteries. *Algorithmic Game Theory*, pages 150–161, 2010.

[6] Amos Fiat, Anna R. Karlin, Elias Koutsoupias, and Angelina Vidali. Approaching utopia: Strong truthfulness and externality-resistant mechanisms. In *Innovations in Theoretical Computer Science (ITCS)*, August 2013. http://arxiv.org/abs/1208.3939v1.

[7] T. Groves. Incentives in Teams. *Econometrica*, 41(4):617–631, 1973.

[8] R.B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[9] N. Nisan and A. Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35(1/2):166–196, 2001.

[10] Noam Nisan. Introduction to mechanism desing (for computer scientists). In Noam Nisan, Tim Roughgarden, Éva Tardos, and Vijay Vazirani, editors, *Algorithmic Game Theory*, chapter 9. Cambridge University Press, 2007.

[11] K. Nissim, R. Smorodinsky, and M. Tennenholtz. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 203–213. ACM, 2012.

[12] William Vickrey. Counterspeculation, auctions and competitive sealed tenders. *The Journal of Finance*, 16(1):8–37, March 1961.

Appendix A

Approaching Utopia: Strong Truthfulness and Externality-Resistant Mechanisms

The paper "Approaching Utopia: Strong Truthfulness and Externality-Resistant Mechanisms" [6] follows.

Approaching Utopia: Strong Truthfulness and Externality-Resistant Mechanisms

Amos Fiat* Anna R. Karlin† Elias Koutsoupias‡ Angelina Vidali§

August 21, 2012

"And verily it is naturally given to all men to esteem their own inventions best."
— Sir Thomas More, in *Utopia*, Book 1, 1516 AD.

Abstract

We introduce and study strongly truthful mechanisms and their applications. We use strongly truthful mechanisms as a tool for implementation in undominated strategies for several problems, including the design of externality resistant auctions and a variant of multi-dimensional scheduling.

1 Introduction

1.1 Externalities

Mechanisms with externalities, and specifically altruism and spite, but also others ("the joy of winning", "malice"), have been studied at length in the literature. Experiments seem to indicate that both altruism and spite have an observable effect, and various theoretical models have been proposed to deal with this issue.

We quote higher authority (Cooper and Fang [15]) in the context of 2nd price auctions: *"We found that small and medium overbids are more likely to occur when bidders perceive their rivals to have similar values, supporting a modified 'joy of winning' hypothesis but large overbids are more likely to occur when bidders believe their opponents to have much higher values, consistent with the 'spite' hypothesis."*

A partial list of (experimental and theoretical) references dealing with externalities is [19, 21, 22, 18, 8, 24, 25, 4, 30, 8, 23, 15, 13, 12, 10, 11]. The questions addressed in previous work primarily deal with the impact of externalities on the equilibria, e.g., observing that externalities such as "the joy or winning" or "spite" lead to overbidding in some auction mechanisms, or that externalities modeled as altruism lead to more-or-less balanced outcomes in the ultimatum game, although neither of these phenomena would be considered "reasonable" if one assumes no externalities. In recent years, the price of anarchy as impacted by such externalities has also been the subject of much research, *e.g.*, malice in congestion games [25, 4, 30].

In this paper we consider a somewhat different goal: we seek to devise mechanisms that *overcome* externalities. As a basic motivating example consider an auction selling a single item. The Vickrey second-price auction is dominant strategy incentive compatible. But, try to imagine that the bidders

*Tel Aviv University. fiat@tau.ac.il
†University of Washington. karlin@cs.washington.edu
‡University of Athens. elias@di.uoa.gr
§University of Vienna. angvid@gmail.com

who lose are spiteful towards the winner (although this is really hard to believe). They may have reason to increase their bid so as to increase the payment by the winning agent.

Even more worrisome — say that the only spiteful losers are those who took part in the various experimental psychology studies cited above, and they did so only so as to mislead the researchers. In fact, we who reside in Utopia will never, ever, encounter spite. This is a fact, but it does *not* imply that everyone *believes* that it is so, it does *not* imply that everyone believes that everyone believes that it is so, it does *not* imply that this is common knowledge. Ergo, just the *concept* of spite (transmitted via the apple from the Garden of Eden), even if in fact there *are no* spiteful bidders, implies that bidders may have an incentive not to bid truthfully in the VCG mechanism.

So, why not define the agent type to include all possible externalities and then run VCG? There are two problems here: (1) It is impossible; payments to one agent impact the utility of another, we are no longer in the quasi-linear setting, (2) Ignoring the former concern (*i.e.*, impossibility), what social welfare are we optimizing? Is it our goal to pander to the spiteful masses? Offer them bread and circuses? Execute the winners during the lunch break of the Gladiatorial games? There are indications from the lives of the Caesars that this may actually maximize (spiteful) social welfare.

So, the very existence of the concept of spite seems to threaten the fundamentals of mechanism design.

To address these issues, we study an alternative utility model: We assume that agents have two utility functions, a *base utility*, and an *externality-modified utility* which is a linear combination of other agent utilities. Variants of this model appear in Ledlard [21], Levine [22], Chen and Kempe [13, 12, 10], and many other papers. The PhD thesis of Chen [10] includes numerous relevant papers.

1.2 Externality Resistant Mechanisms

We present a new type of private value mechanism, rVCG. Assume that it is common knowledge that no one is willing to lose more than (say) $\gamma = 5$ cents so as to increase another's payment by \$1. Now:

1. Agents using the rVCG mechanism are sure that the following two values are approximately equal:

    - The utility they obtain under rVCG, in an imperfect world, where externalities are real, and demons roam the earth.
    - The utility they would have obtained under VCG, in an imaginary, Utopian world, where externalities did not exist. (See Theorem 3.1).

    I.e., given a bound, γ, on the altruism/spite, the rVCG mechanism approximates Utopia, as promised in the title.[1]

2. On the other hand, irrespective of how infinitesimally small $\gamma > 0$ may be, a losing bidder in a second-price auction, may, out of spite, even infintesimally small spite, reduce the winner's profit to zero. (This holds in the more general VCG mechanism as well).

1.3 Strongly Truthful Mechanisms

To achieve externality resistant mechanisms we make use of strongly truthful mechanisms. These are mechanisms where it is not only a weakly dominant strategy to be truthful but where one gets punished for lying. The goal in the design of strongly truthful mechanisms is to increase

[1] Admittedly, the bound γ has to be very small in order to truly approximate Utopia.

the punishment as much as possible. Strongly truthful mechanisms are related to strongly convex mechanisms, analogous to the connection between truthful mechanisms and convex utility functions, (see, *e.g.*, Archer and Kleinberg, [2, 1]).

For bounded domains, we give (optimal) strongly truthful mechanisms, in this case, the punishment for the lie $\tilde{v} = v + \delta$ is $O(\delta^2)$.

For unbounded domains, we give a mechanism that is *relatively strongly truthful* where the lie is measured as a fraction of the truth, and the punishment for the lie $\tilde{v} = (1+\alpha)v$, where $\alpha \in \Theta(1)$, is $v/\log^{1+\epsilon} v$.

Strongly truthful mechanisms can also be used in mechanisms for multi-dimensional problems such as makespan minimization for unrelated machines, see below.

This idea of combining multiple mechanisms to boost truthfulness appears in [28], where it is used to derive truthful mechanisms for some problems via differential privacy. It also appears implicitly in the context of scoring rules [9, 6], and in the related responsive lotteries [16] so as to determine the true utility of an outcome. However, we are unaware of previous attempts to quantify the quality of such devices, nor are we aware of other attempts to apply them towards externality resistance or for multidimensional problems.

In the appendix we describe transformations between strongly truthful mechanisms and proper scoring rules. This automatically implies transformations between strongly truthful mechanisms, market scoring rules, responsive lotteries, and market maker pricing algorithms to provide liquidity for prediction markets [17, 31, 14].

1.4 The Solution Concept

Adapting a solution concept from Babaioff, Lavi and Pavlov, [5], from approximation problems to arbitrary predicates, we say that a mechanism M is an algorithmic implementation of a predicate P in undominated strategies, if, for all agents i, there exists a set of strategies, D_i, such that

1. The output of M satisfies P, for any combination of strategies from $\prod_j D_j$, and,
2. For all i, for any agent i strategy, $s \notin D_i$, there exists some strategy $s' \in D_i$ that is strictly better for agent i than strategy s, irrespective of what strategies are chosen by the other agents.

I.e., predicate P is implemented by a mechanism in undominated strategies, if, in the game defined by the mechanism, and as long as no agent chooses a strategy that is obviously dominated (for arbitrary assumptions about the types of other agents, *e.g.*, values, bids, externalities), predicate P holds for the outcome of the mechanism.

In the context of externality resistant auctions, as long as agents do not bid stupidly (do not use a strategy that is obviously dominated), externality resistance holds.

In fact, any strategy that entails bidding "too far away" from the truth is dominated by bidding truthfully, where "too far away" for agent i is a function of her own externalities γ_{ij} (see Section 3 for a definition of these externalities). Moreover, agents can efficiently determine that bidding far from the truth is dominated by truthful bidding. Thus, D_i is contained in the set of all bids whose distance from the truth is not too big. Note that we don't make any claim on the precise strategies that will be adapted by the agents.

1.5 Other Applications

We can also use strongly truthful mechanisms to achieve goals such as minimizing the makespan in a multi dimensional machine scheduling problem, the infamous Nisan-Ronen problem, see [27, 20, 3].

This magic is achieved by changing the problem, and allowing one to repeatedly assign the same job to a machine. So, choosing to verify the truthfullness of the agent types can be done by choosing at random, with some small probability, a target agent, and using strongly truthful mechanisms to punish the agent for misrepresentation of his type.

Given a sufficiently large punishment, all agents will have incentive to stick close to the truth. So, with high probability, the mechanism will achieve a close approximation to the minimum makespan in undominated strategies..

This is quite general and can be used in other multi-dimensional settings where one can boost truth extraction by repetition.

2 Strongly Truthful Mechanisms

A key ingredient in our constructions is the notion of a *strongly truthful mechanism*. In this section, we define strongly truthful mechanisms for single dimensional problems and one agent. As discussed below, these definitions and results extend to multi-dimensional and multi-agent settings.

Consider a single dimensional agent with private value (type) v for receiving a good or service. A direct revelation mechanism takes as input some (possibly false) value, $\tilde{v}$, computes a payment, $p(\tilde{v})$, and allocates the good to the agent with probability $a(\tilde{v})$.

The standard quasilinear utility of an agent whose true value is v, but reports value $\tilde{v}$ (possibly different from v), is denoted by

$$u_v(\tilde{v}) = v \cdot a(\tilde{v}) - p(\tilde{v}). \tag{1}$$

We also define

$$u(v) = u_v(v),$$

i.e., the utility to the agent with value v when truthfully reporting $\tilde{v} = v$.

In this setting, it follows from Myerson [26], that a mechanism is truthful in expectation if and only if

- the allocation probability function $a(v)$ is monotone nondecreasing, and
- the payment function is

$$p(v) = va(v) - \int_0^v a(x)dx + p(0),$$

  for some constant $p(0)$. (We will take $p(0) = 0$ herein).

It follows from the above and from Equation 1 that

$$\begin{aligned} u_v(\tilde{v}) &= v \cdot a(\tilde{v}) - \tilde{v} \cdot a(\tilde{v}) + \int_0^{\tilde{v}} a(x)dx \\ &= (v - \tilde{v}) \cdot a(\tilde{v}) + \int_0^{\tilde{v}} a(x)dx. \end{aligned}$$

Thus, for truthful in expectation mechanisms, it must be that

1. The utility function $u(v)$ is convex (the integral of a nondecreasing function).
2. The allocation function $a(v) = u'(v)$ (where u is differentiable).
3. Any convex function $u(v)$ whose subgradient, $u'(v)$, lies in the range $[0, 1]$, can be interpreted as the utility function for an associated truthful in expectation mechanism.

4. Ergo, if restricting oneself to truthful in expectation mechanisms, one can describe a mechanism using utility functions or allocation functions interchangeably. (Up to additive constants).

We seek to strengthen the notion of truthfulness in expectation so that the greater the deviation from the truth, the greater the loss in utility.

To this end, we define c-strongly truthful mechanisms as follows:

Definition 2.1. *A mechanism with utility function u is called c-strongly truthful if for every v and $\tilde{v}$:*

$$u_v(v) - u_v(\tilde{v}) \geq \frac{1}{2} c \, |\tilde{v} - v|^2. \tag{2}$$

This definition enables us to extend the connection between truthfulness and convexity to strongly truthful mechanisms. For this, recall the standard notion of strong convexity. For a differentiable function $f(x)$, convexity is equivalent to:

$$\forall x, x' \quad f(x) - f(x') \geq f'(x') \cdot (x - x').$$

The following notion is also standard [7]:

Definition 2.2. *Let $m \geq 0$. A function f is called m-strongly convex if and only if for every x, x':*

$$f(x) - f(x') \geq f'(x') \cdot (x - x') + \frac{1}{2} m \, |x - x'|^2 \tag{3}$$

By defining strong truthfulness as in Equation (2) the following proposition holds:

Lemma 2.3. *A mechanism with utility function $u(v)$ is m-strongly truthful if and only if $u(v)$ is m-strongly convex.*

Proof. Applying equation (1) to $u_v(\tilde{v})$ and $u_{\tilde{v}}(\tilde{v})$, we get

$$u_{\tilde{v}}(\tilde{v}) - u_v(\tilde{v}) = u'(\tilde{v}) \cdot (\tilde{v} - v).$$

It follows that

$$u_v(v) - u_v(\tilde{v}) = u_v(v) - u_{\tilde{v}}(\tilde{v}) + u'(\tilde{v}) \cdot (\tilde{v} - v) = u(v) - u(\tilde{v}) + u'(\tilde{v}) \cdot (\tilde{v} - v) \tag{4}$$

Since by definition, the mechanism is m-strongly truthful if and only if $u_v(v) - u_v(\tilde{v}) \geq \frac{1}{2} m \, |\tilde{v} - v|^2$, we derive that the mechanism is m-strongly truthful if and only if $u(v) - u(\tilde{v}) + u'(\tilde{v}) \cdot (\tilde{v} - v) \geq \frac{1}{2} m \, |\tilde{v} - v|^2$, which is precisely the definition that $u(v)$ is m-strongly convex. ☐

Remark: All of the definitions in this section extend naturally to multi-dimensional agents. Indeed, the three equivalent definitions of a doubly differentiable function being convex (the standard one, cycle monotonicity, and the Hessian being positive semidefinite) have analogues when discussing truthful multidimensional mechanisms over convex domains [7, 29]. Similarly, the equivalent notions of strong-convexity and strong truthfulness extend mutatis mutandis.

It follows from the above theorem that the question of finding the strongest truthful mechanism is an extremal question about strongly convex functions whose partial derivatives satisfy appropriate constraints that capture the constraints of the allocation probabilities (for example, for the single item case the constraint is the derivative of the utility is in $[0, 1]$).

2.1 Strongly Truthful Mechanisms for Single Agent, Single Item Auctions

Consider the case in which we want to find the strongest truthful mechanism for a single player and one item. (We will use this in the next section.) To start, assume that the agent's value is bounded: $v \in [L, H]$. For this case, we define the *linear mechanism*:

Definition 2.4. *The linear mechanism for the single player/single item setting has allocation rule $a(v) = (v - L)/(H - L)$, and applies when the player's value is known to be in the range $[L, H]$.*

Theorem 2.5. *The linear mechanism for a player whose value v satisfies $v \in [L, H]$ is a $1/(H-L)$-strongly truthful mechanism. No other mechanism is m-strongly truthful for $m \geq 1/(H - L)$.*

Proof. It is straightforward to check that for the linear mechanism, the utility function $u(v)$ is $\frac{(v-L)^2}{2(H-L)}$. We can directly verify that Equation (3) in the definition of strong convexity holds with equality for all v, with $m = 1/(H - L)$. Indeed, we derive the following equivalences

$$u(z) - u(y) = u'(y) \cdot (z - y) + \frac{1}{2} m \, |z - y|^2$$
$$\frac{(z - L)^2}{2(H - L)} - \frac{(y - L)^2}{2(H - L)} = \frac{y - L}{H - L}(z - y) + \frac{1}{2}\frac{1}{H - L}(z - y)^2$$
$$\frac{(z - y)(z + y - 2L)}{2(H - L)} = \frac{(z - y)(2(y - L) + (z - y))}{2(H - L)};$$

the last equality clearly holds.

We now show that this is the strongest truthful mechanism. From the definition of strong convexity for the extreme values of the domain, we get

$$u(H) - u(L) \geq u'(L)(H - L) + \frac{1}{2} m(H - L)^2$$
$$u(L) - u(H) \geq u'(H)(L - H) + \frac{1}{2} m(L - H)^2$$

Adding these two, we get that

$$(H - L)(u'(H) - u'(L)) \geq m(H - L)^2$$

Since $u'(L)$, and $u'(H)$ are in $[0, 1]$ (they represent allocations), we get that $m \leq 1/(H - L)$. ☐

Remarks:

- There is a direct connection between single-agent truthful mechanisms and scoring rules (see e.g., [6]). Indeed, one can define a notion of *strongly proper scoring rules* that is analogous to a strongly truthful mechanism. We note that the mechanism just described is in fact the well-known quadratic scoring rule.

- Definition 2.4 and Theorem 2.5 can easily be generalized to the case of one player and many items with additive valuations. In this case the utility is $u(v) = \sum_{j=1}^{m} \frac{(v_j - L)^2}{2(H-L)}$, for which $m = \frac{1}{n(H-L)}$.

2.2 Relative strong truthfulness

If we want to consider unbounded domains, it follows from Theorem 2.5 that no m-strongly truthful mechanism exists with $m > 0$.

For such domains, it may be useful to define a notion of relative strong truthfulness as follows:

Definition 2.6. *We say a mechanism M is $f(v, \alpha)$-relatively truthful if, for all $\tilde{v}$ such that $\tilde{v} \notin [v(1-\alpha), v(1+\alpha)]$*

$$\frac{u_v(v) - u_v(\tilde{v})}{u_v(v)} \geq f(v, \alpha).$$

For example, it is easy to show that the single agent mechanism with allocation rule $a(v) = 1 - \frac{1}{\ln v} + \frac{1}{\ln^2 v}$ (and payment $p(v) = \frac{v}{\ln^2(v)}$) satisfies $f(v, \alpha) = \Omega(\frac{\alpha^2}{\log^2 v})$. Slightly better mechanisms that approach $f(v, \alpha) = \Omega(\frac{\alpha^2}{\log v})$ exist[2].

3 Externality Resistant Auctions

In this section, we consider how strongly truthful, or truth-extraction mechanisms can be used to help cope with spiteful or altruistic bidders. Our goal is to ensure that a bidder participating in, say, an auction for a single item, does not need to worry about her competitor purposely bidding high just so as to make her pay a lot.

We consider the setting where an auctioneer wishes to maximize social welfare, and each agent has a value v_i for being one of the winners in the auction. Of course, in the standard version of this setting, the mechanism of choice would be the VCG mechanism.

As we have already discussed however, the VCG mechanism is entirely vulnerable to spiteful agents. Before explaining the alternative we propose, we define a utility model for externalities that captures precisely what we mean when we speak about spiteful and altruistic agents.

In the *externality-modified* setting, agent i's type t_i consists of

- v_i, her value for service; and
- a set of externality parameters γ_{ij} for all $j \neq i$. Intuitively, γ_{ij} represents how much agent i cares about the utility of agent j. A large, negative value means that i is significantly motivated by the desire to decrease agent j's utility, whereas a large, positive value means that i seeks to increase agent j's utility. A value of zero means that i is indifferent towards j.

Let $\mathcal{M}$ be an arbitrary mechanism for the single-parameter allocation problem under consideration. The mechanism takes as input a bid b_i from each agent (which is equal to v_i if the mechanism is truthful) and produces as output an allocation $\mathbf{x}$, where x_i is the probability that agent i receives service, and payments $\mathbf{p}$, with p_i the expected payment by agent i. Note that both $x_i = x_i(\mathbf{b})$ and $p_i = p_i(\mathbf{b})$ are functions of the bids. The allocation selected must satisfy the feasibility constraints of the setting, however, we do assume, that having only a single arbitrary agent receive service is feasible.

[2]The following sequence of mechanisms are defined for every $k > 1$ and approach $f(v, \alpha) = \Omega(\frac{\alpha^2}{\log v})$:

$$p(v) = \frac{v}{\ln^k v} \qquad a(v) = 1 - \frac{1}{(k-1)\ln^{k-1} v} + \frac{1}{\ln^k v}$$

$$p(v) = \frac{v}{\ln v \ln^k \ln v} \qquad a(v) = 1 - \frac{1}{(k-1)\ln v \ln^{k-1} \ln v} + \frac{1}{\ln v \ln^k \ln v}$$

and so on.

Given bids $\mathbf{b}_{-i}$ of all players except player i, the *base (standard) utility* of agent i, when her type is $t_i = (v_i, \{\gamma_{i1}, \ldots, \gamma_{in}\})$ and her bid is b_i, is denoted by $u^{\mathcal{M}}_{v_i}(b_i, \mathbf{b}_{-i})$ and is defined as

$$u^{\mathcal{M}}_{v_i}(b_i, \mathbf{b}_{-i}) = v_i\, x_i(\mathbf{b}) - p_i(\mathbf{b}). \tag{5}$$

Notice that this utility depends only on v_i and not on the rest of agent i's private information (agent i's externality parameters γ_{ij}). This is why we subscript the utility by v_i instead of t_i.

We define the *externality-modified* utility $\widehat{u}^{\mathcal{M}}_{t_i}$ of agent i when the types of the agents are $\mathbf{t}$ and the bids are $\mathbf{b}$ as

$$\widehat{u}^{\mathcal{M}}_{t_i}(\mathbf{b}, \mathbf{t}_{-i}) = u^{\mathcal{M}}_{v_i}(b_i, \mathbf{b}_{-i}) + \sum_{j \neq i} \gamma_{ij}\, u^{\mathcal{M}}_{v_j}(b_j, \mathbf{b}_{-j}). \tag{6}$$

This model (and variants thereof) have been used previously in several papers, e.g. [22, 10].

Note that

- Because the externality-modified utility defined above depends, not only on the bids (or actions) of other agents, b_{-i}, but also on their types, we add the t_{-i} as an argument to the utility function, which is atypical. (Of course the only part of t_{-i} the utility depends on is v_{-i}.)
- The value of t_{-i} is, in general, unknown to agent i, so agent i will be, in general, unable to compute her externality modified utility $\widehat{u}^{\mathcal{M}}_{t_i}(\mathbf{b}, \mathbf{t}_{-i})$.
- We will be particularly interested in cases where the mechanism $\mathcal{M}$ that is being run is VCG, and then use $u^{\mathrm{VCG}}_{v_i}(\mathbf{b})$ to denote the standard utility of agent i when her value is v_i, the reported bids are $\mathbf{b}$ and the mechanism being run is VCG.

Our goal is to design a mechanism that is *externality-resistant*, in the following sense:

- The mechanism approximately maximizes social welfare.
- Despite the fact that each agent bids to maximize their externality-modified utility, each agent ends up with a base utility that is approximately what it would have been had all agents bid so as to maximize their base utility. Thus, non-spiteful agents are not harmed by the presence of spiteful agents. Furthermore, the auctioneer's revenue is not harmed by the presence of altruistic agents.

An immediate difficulty that arises is the fact that our utility model is non-quasi linear. Our approach is to consider a weaker solution concept, *implementation in undominated strategies,* formally defined as follows.

In a game of incomplete information, a strategy for an agent is a function mapping types to actions. We say that a strategy s'_i for agent i is *dominated* by strategy s_i if for all types t_i for agent i, and for all possible types t_{-i} and all possible actions

$$b_{-i} = s_{-i}(t_{-i})$$

of the other agents, the utility of agent i satisfies:

$$u_{t_i}(s_i(t_i), b_{-i}, t_{-i}) \geq u_{t_i}(s'_i(t_i), b_{-i}, t_{-i}).$$

A strategy s_i for agent i is *undominated* if it is not dominated.

We say that a mechanism implements a predicate P in undominated strategies, if whenever agents are limited to playing undominated strategies, it must be that predicate P holds.

We consider the following simple variant of VCG, which we call *externality-resistant* VCG or rVCG , for short. The rVCG mechanism, with n agents participating, is parameterized by a value δ, $0 \le \delta \le 1$, and works as follows:

- Ask the n agents for their values/bids.
- With probability δ/n single out the i-th agent and run the truth extraction mechanism (denoted by TE) on him.
- With probability $1-\delta$, run VCG.

Our main theorem is the following:

Theorem 3.1. *Consider any single-parameter allocation setting in which the n agents true values v_i are all in the range $[0,1]$. Let γ denote $\max_{ij}\gamma_{ij}$. Then, for any n, δ, ϵ, and γ_{ij} such that*

$$\gamma < \frac{\epsilon\,\delta}{8\,(1-\delta)^2\,n^3},$$

mechanism rVCG *above implements the following predicate in undominated strategies:*

For all agents i, and for all types $\mathbf{t}$, the base utility obtained under the rVCG *mechanism is close to the base utility obtained by agent i when all agents bid truthfully under the "standard" VCG mechanism. Specifically, for* $\mathbf{b}$ *undominated,*

$$u_{v_i}^{\mathrm{rVCG}}(\mathbf{b}) \ge (1-\delta)u_{v_i}^{\mathrm{VCG}}(\mathbf{v}) - \epsilon.$$

In the proof below, we use the following notation and definitions:

- For any set of bids $\mathbf{b}$, let $\mathrm{MSW}(\mathbf{b})$ denote the maximum social welfare achievable with respect to the bids $\mathbf{b}$, i.e.
$$\mathrm{MSW}(\mathbf{b}) = \max_a \sum_j b_j(a).$$
- We define $\mathrm{MSW}_{v_i}(b_i, \mathbf{b}_{-i})$ to be the maximum social welfare *experienced* by agent i, when agent i bids b_i, whereas her true value is v_i, and all other agents bid $\mathbf{b}_{-i}$. Thus,
$$\mathrm{MSW}_{v_i}(b_i, \mathbf{b}_{-i}) = v_i(a^*) + \sum_{j\ne i} b_j(a^*), \text{ where } \ a^* = \mathrm{argmax}_a \sum_j b_j(a).$$
- When agent i bids b_i, her true value is v_i, all agents but i bid $\mathbf{b}_{-i}$, then, the utility of agent i under VCG with Clarke Pivot Payments is
$$u_{v_i}^{\mathrm{VCG}}(\mathbf{b}) = \mathrm{MSW}_{v_i}(b_i, \mathbf{b}_{-i}) - \mathrm{MSW}(\mathbf{b}_{-i}).$$

We can now turn to the proof of the theorem.

Proof. We assume that agents would like to maximize their externality-modified utility:

$$\widehat{u}_{t_i}^{\mathrm{rVCG}}(\mathbf{b}, \mathbf{t}_{-i}) = (1-\delta)\widehat{u}_{t_i}^{\mathrm{VCG}}(\mathbf{b}, \mathbf{t}_{-i}) + \frac{\delta}{n}\left(u_{v_i}^{\mathrm{TE}}(b_i) + \sum_{j\ne i}\gamma_{ij}u_{v_j}^{\mathrm{TE}}(b_j)\right), \tag{7}$$

where

$$\widehat{u}_{t_i}^{\mathrm{VCG}}(\mathbf{b}, \mathbf{t}_{-i}) = u_{v_i}^{\mathrm{VCG}}(\mathbf{b}) + \sum_{j \neq i} \gamma_{ij} \left(u_{v_j}^{\mathrm{VCG}}(\mathbf{b}) \right). \tag{8}$$

We say an agent is *standard* if $\gamma_{ij} = 0$ for all j. Such an agent doesn't care about the utility of others. For standard agents, the base utility and the externality-modified utility are the same. In addition, because VCG is dominant strategy truthful, it is a dominant strategy for the standard agents to bid truthfully.

So, we only need to understand how non-standard agents will bid. To this end, fix an agent i of type t_i and the bids $\mathbf{b}_{-i}$ and types $\mathbf{t}_{-i}$ of the other agents.

Suppose that the VCG part of rVCG is executed. Then agent i's externality-modified utility $\widehat{u}_{t_i}^{\mathrm{VCG}}(b_i, \mathbf{b}_{-i}, \mathbf{t}_{-i})$ is defined by equation (8). We have

$$u_{v_j}^{\mathrm{VCG}}(\mathbf{b}) = \mathrm{MSW}_{v_j}(b_j, \mathbf{b}_{-j}) - \mathrm{MSW}(\mathbf{b}_{-j})$$

and

$$u_{v_j}^{\mathrm{VCG}}(b_j, v_i, \mathbf{b}_{-ij}) = \mathrm{MSW}_{v_j}(b_j, v_i, \mathbf{b}_{-ij}) - \mathrm{MSW}(v_i, \mathbf{b}_{-i,j}).$$

Thus, the difference between agent i's externality modified utility when agent i bids b_i and when agent i bids v_i is given by

$$\begin{aligned}
\widehat{u}_{t_i}^{\mathrm{VCG}}(b_i, \mathbf{b}_{-i}, \mathbf{t}_{-i}) - \widehat{u}_{t_i}^{\mathrm{VCG}}(v_i, \mathbf{b}_{-i}, \mathbf{t}_{-i}) &= \mathrm{MSW}_{v_i}(b_i, \mathbf{b}_{-i}) - \mathrm{MSW}(v_i, \mathbf{b}_{-i}) \\
&\quad + \sum_{j \neq i} \gamma_{ij} \left(\mathrm{MSW}_{v_j}(b_j, \mathbf{b}_{-j}) - \mathrm{MSW}_{v_j}(b_j, v_i, \mathbf{b}_{-ij}) \right) \\
&\quad + \sum_{j \neq i} \gamma_{ij} \left(\mathrm{MSW}(v_i, \mathbf{b}_{-i,j}) - \mathrm{MSW}(b_i, \mathbf{b}_{-i,j}) \right) \\
&\leq 2 \sum_{j \neq i} \gamma_{ij} \eta_i \\
&\leq 2(n-1)\gamma_i \eta_i,
\end{aligned} \tag{9}$$

where $|b_i - v_i| = \eta_i$ and $\gamma_i = \max_j \gamma_{ij}$.

On the other hand,

$$\widehat{u}_{v_i}^{\mathrm{TE}}(v_i) - \widehat{u}_{v_i}^{\mathrm{TE}}(b_i) \geq \frac{1}{2} m(b_i) |b_i - v_i|^2 \geq \frac{1}{2} \eta_i^2, \tag{10}$$

assuming valuations in the range $[0, 1]$ and the use of the linear TE algorithm. Combining inequalities (9) and (10), we have

$$\widehat{u}_{t_i}^{\mathrm{rVCG}}(b_i, \mathbf{b}_{-i}) - \widehat{u}_{t_i}^{\mathrm{rVCG}}(v_i, \mathbf{b}_{-i}) \leq (1-\delta) 2(n-1)\gamma \eta_i - \frac{\delta}{2n} \eta_i^2.$$

Consider any bid b_i for which $\eta_i = |b_i - v_i|$ has

$$(1-\delta) 2(n-1)\gamma \eta_i - \frac{\delta}{2n} \eta_i^2 < 0.$$

Then the strategy of bidding truthfully dominates the strategy of bidding this b_i. Thus, for all undominated strategies, it must be that agent i bids a value b_i which satisfies:

$$0 \leq \widehat{u}_{t_i}^{\mathrm{rVCG}}(b_i, \mathbf{b}_{-i}, \mathbf{t}_{-i}) - \widehat{u}_{t_i}^{\mathrm{rVCG}}(v_i, \mathbf{b}_{-i}, \mathbf{t}_{-i}),$$

implying that she will choose η_i so that

$$0 \leq (1-\delta)2(n-1)\gamma\eta_i - \frac{\delta}{2n}\eta_i^2,$$

and thus

$$\eta_i \leq \frac{4(1-\delta)}{\delta} n^2 \gamma. \tag{11}$$

In other words, *lying about v_i by more than the right-hand side of Equation (11) is a strategy dominated by the truth-telling strategy.*

From this we can conclude that for any player ℓ who participates in rVCG, if all agents play undominated strategies, then:

$$\begin{aligned}
u_{v_\ell}^{\mathrm{rVCG}}(b_\ell, \mathbf{b}_{-\ell}) &= (1-\delta) u_{v_\ell}^{\mathrm{VCG}}(b_\ell, \mathbf{b}_{-\ell}) + \frac{\delta}{n} u_{v_\ell}^{\mathrm{TE}}(b_\ell) \\
&= (1-\delta)\left(u_{v_\ell}^{\mathrm{VCG}}(v_\ell, \mathbf{v}_{-\ell}) - \left(u_{v_\ell}^{\mathrm{VCG}}(v_\ell, \mathbf{v}_{-\ell}) - u_{v_\ell}^{\mathrm{VCG}}(b_\ell, \mathbf{b}_{-\ell})\right)\right) \\
&\quad + \frac{\delta}{n} u_{v_\ell}^{\mathrm{TE}}(b_\ell) \\
&\geq (1-\delta)\left(u_{v_\ell}^{\mathrm{VCG}}(v_\ell, \mathbf{v}_{-\ell}) - 2n\eta\right) \\
&\geq (1-\delta) u_{v_\ell}^{\mathrm{VCG}}(v_\ell, \mathbf{v}_{-\ell}) - \frac{8(1-\delta)^2}{\delta} n^3 \gamma
\end{aligned} \tag{12}$$

where $\eta = \max_i \eta_i$.

□

The following two corollaries are immediate:

Corollary 3.2. *When* rVCG *is used and all players play undominated strategies, the social welfare of the outcome a^* selected satisfies*

$$\sum_i v_i(a^*) \geq \mathrm{MSW}(\mathbf{v}) - n\eta.$$

Corollary 3.3. *When* rVCG *is used and all players play undominated strategies, the profit of the auctioneer is at least his profit from running VCG with truthful players minus $2n\eta$.*

4 Discussion

In this paper, we have introduced a number of concepts and taken first steps towards understanding and applying these concepts. Clearly though, we have only scratched the surface.

For example, the basic tool of strongly truthful mechanisms may have some potential, but there is clearly much left to be understood. What is the right way to define strong truthfulness? What mechanisms achieve optimal relative strong truthfulness? What is the tradeoff between the "strength" of the truthfulness and the social welfare that can be achieved?

We explored a new utility model for externalities in the context of mechanism design and sought to design mechanisms that protect agents from these externalities. Our mechanism has some externality-resistance, however, the externality parameters have to be extremely small in order for our mechanism to be effective. Is it possible to do better? More concretely, our mechanism tolerates externality parameters of value $\gamma = O(1/n^3)$. Are there mechanisms that tolerate higher values of γ? In the opposite direction, can we show a bound on the maximum γ?

What happens if we use a different solution concept? Also, while we chose to optimize for "base utility", that is not the only goal one might consider. One could optimize for social welfare with respect to the externality modified utilities. To what extent is this possible? Is this a reasonable goal? *I.e.*, should the goal of the mechanism be to encourage spite?

References

[1] Aaron Archer and Robert Kleinberg. Characterizing truthful mechanisms with convex type spaces. *SIGecom Exchanges*, 7(3), 2008.

[2] Aaron Archer and Robert Kleinberg. Truthful germs are contagious: a local to global characterization of truthfulness. In Lance Fortnow, John Riedl, and Tuomas Sandholm, editors, *ACM Conference on Electronic Commerce*, pages 21–30. ACM, 2008.

[3] Itai Ashlagi, Shahar Dobzinski, and Ron Lavi. An optimal lower bound for anonymous scheduling mechanisms. In *ACM Conference on Electronic Commerce (EC)*, pages 169–176, 2009.

[4] Moshe Babaioff, Robert Kleinberg, and Christos H. Papadimitriou. Congestion games with malicious players. In *Proceedings of the 8th ACM conference on Electronic commerce*, EC '07, pages 103–112, New York, NY, USA, 2007. ACM.

[5] Moshe Babaioff, Ron Lavi, and Elan Pavlov. Single-value combinatorial auctions and implementation in undominated strategies. In *SODA*, pages 1054–1063. ACM Press, 2006.

[6] J. Eric Bickel. Some comparisons among quadratic, spherical, and logarithmic scoring rules. *Decision Analysis*, 4(2):49–65, 2007.

[7] S.P. Boyd and L. Vandenberghe. *Convex optimization*. Cambridge Univ Pr, 2004.

[8] F. Brandt, T. Sandholm, and Y. Shoham. Spiteful bidding in sealed-bid auctions. In *Proceedings of IJCAI'07*, 2007.

[9] Glenn W. Brier. Verification of forecasts expressed in terms of probability. *Mon. Wea. Rev.*, 78:1–3, 1950.

[10] Po-An Chen. The effects of altruism and spite on games. *PhD Thesis, University of Southern California*, 59(4):757–75, October 2011.

[11] Po-An Chen, Bart de Keijzer, David Kempe, and Guido Schäfer. The robust price of anarchy of altruistic games. In Ning Chen, Edith Elkind, and Elias Koutsoupias, editors, *WINE*, volume 7090 of *Lecture Notes in Computer Science*, pages 383–390. Springer, 2011.

[12] Po-An Chen and David Kempe. Altruism, selfishness, and spite in traffic routing. In *Proceedings of the 9th ACM conference on Electronic commerce*, EC '08, pages 140–149, New York, NY, USA, 2008. ACM.

[13] Po-An Chen and David Kempe. Bayesian auctions with friends and foes. In Marios Mavronicolas and Vicky Papadopoulou, editors, *Algorithmic Game Theory*, volume 5814 of *Lecture Notes in Computer Science*, pages 335–346. Springer Berlin / Heidelberg, 2009.

[14] Yiling Chen. A utility framework for bounded-loss market makers. In *In Proceedings of the 23rd Conference on Uncertainty in Artificial Intelligence*, pages 49–56, 2007.

[15] David J. Cooper and Hanming Fang. Understanding overbidding in second price auctions: An experimental study. *The Economic Journal*, 118(532):1572–1595, 2008.

[16] Uriel Feige and Moshe Tennenholtz. Responsive lotteries. In Spyros C. Kontogiannis, Elias Koutsoupias, and Paul G. Spirakis, editors, *SAGT*, volume 6386 of *Lecture Notes in Computer Science*, pages 150–161. Springer, 2010.

[17] Robin Hanson. Logarithmic market scoring rules for modular combinatorial information aggregation. *The Journal of Prediction Markets*, 1(1):3–15, 2007.

[18] Philippe Jehiel, Benny Moldovanu, and Ennio Stacchetti. Multidimensional mechanism design for auctions with externalities. *Journal of Economic Theory*, 85(2):258–293, 1999.

[19] John H. Kagel, Ronald M. Harstad, and Dan Levin. Information impact and allocation rules in auctions with affiliated private values: A laboratory study. *Econometrica*, 55(6):pp. 1275–1304, 1987.

[20] Ron Lavi and Chaitanya Swamy. Truthful mechanism design for multi-dimensional scheduling via cycle monotonicity. In *ACM Conference on Electronic Commerce (EC)*, pages 252–261, 2007.

[21] John O. Ledyard. Public goods: A survey of experimental research. Public Economics 9405003, EconWPA, May 1994.

[22] David K. Levine. Modeling altruism and spitefulness in experiments. *Review of Economic Dynamics*, 1(3):593 – 622, 1998.

[23] Emiel Maasland and Sander Onderstal. Auctions with financial externalities. *Economic Theory*, 32(3):551–574, September 2007.

[24] John Morgan, Ken Steiglitz, and George Reis. The spite motive and equilibrium behavior in auctions. *The B.E. Journal of Economic Analysis & Policy*, 0(1):5, 2003.

[25] Thomas Moscibroda, Stefan Schmid, and Roger Wattenhofer. The price of malice: A game-theoretic framework for malicious behavior in distributed systems. *Internet Mathematics*, 6(2):125–155, 2009.

[26] Roger B. Myerson. Optimal auction design. *Mathematics of Operations Research*, 6(1):58–73, 1981.

[27] Noam Nisan and Amir Ronen. Algorithmic mechanism design. *Games and Economic Behavior*, 35:166–196, 2001.

[28] Kobbi Nissim, Rann Smorodinsky, and Moshe Tennenholtz. Approximately optimal mechanism design via differential privacy. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS '12, pages 203–213, New York, NY, USA, 2012. ACM.

[29] R.T. Rockafellar. *Convex analysis*, volume 28. Princeton Univ Pr, 1997.

[30] Aaron Roth. The price of malice in linear congestion games. In Christos H. Papadimitriou and Shuzhong Zhang, editors, *WINE*, volume 5385 of *Lecture Notes in Computer Science*, pages 118–125. Springer, 2008.

[31] Justin Wolfers and Eric Zitzewitz. Prediction markets. Working Paper 10504, National Bureau of Economic Research, May 2004.

A Strong Truthfulness and Scoring Rules

In this appendix, we develop the connection between strongly truthful single-agent mechanisms and "strongly proper" scoring rules [9, 6]. As mentioned above, this immediately relates strongly truthful mechanisms to a host of seemingly unrelated problems.

A.1 The setting

We consider the following setting:

- There is a set of $n+1$ events, (call them events 0 through n) one of which will happen.
- The agent (forecaster) has a belief vector $\mathbf{p}$ as to which event will happen, where p_i is the probability that event i happens. $p_0 = 1 - \sum_{1 \le i \le n} p_i$.
- The mechanism takes as input a postulated belief vector $\tilde{\mathbf{p}}$, and uses a scoring rule to determine the "payments" or "scores". Specifically, the scoring rule says for each outcome i, the "payment" or "score" the agent gets is $s_i(\tilde{\mathbf{p}})$.
- The agent proposes $\tilde{\mathbf{p}}$ and obtains utility
$$\sum_{0 \le i \le n} p_i s_i(\tilde{\mathbf{p}}).$$
- The scoring rule is strictly proper if reporting $\tilde{\mathbf{p}} = \mathbf{p}$ strictly maximizes his utility.

A.2 Translating Mechanisms to Scoring Rules

Let M be a mechanism that takes as input an agent's valuations $x_1, \ldots, x_n$ for n alternatives. We assume $x_i \ge 0$ for all i and that $\sum_i x_i \le 1$. The mechanism has allocation probabilities $a_i(\mathbf{x})$ and a payment rule $P(\mathbf{x})$.

We convert this to a scoring rule $S(M)$ as follows: Given vector $\mathbf{p}$ representing the probabilities $p_1, \ldots, p_n$ of outcomes (with $p_0 = 1 - \sum_i p_i$), let $s_i(\mathbf{p}) = a_i(\mathbf{p}) - P(\mathbf{p})$, and let $s_0(\mathbf{p}) = -P(\mathbf{p})$.

Proposition 1. *If M is strictly truthful (i.e., it is strictly optimal to be truthful), then $S(M)$ is strictly proper.*

Proof. By definition, the payoff to the agent when using the scoring rule and reporting $\tilde{\mathbf{p}}$ is

$$\sum_{1 \le i \le n} p_i s_i(\tilde{\mathbf{p}}) + \left(1 - \sum_{1 \le i \le n} p_i\right) s_0(\tilde{\mathbf{p}})$$

This is the same as

$$\sum_{1 \le i \le n} p_i \left(a_i(\tilde{\mathbf{p}}) - P(\tilde{\mathbf{p}})\right) - \left(1 - \sum_{1 \le i \le n} p_i\right) P(\tilde{\mathbf{p}}) = \sum_i p_i a_i(\tilde{\mathbf{p}}) - P(\tilde{\mathbf{p}}).$$

and thus the incentives for the scoring rule are identical to the incentives for the mechanism. $\square$

A.3 Translating Scoring Rules to Mechanisms

Let S be a non-trivial scoring rule. We assume that the s_i's are of bounded absolute value. We show how to convert this to a mechanism:

Define the constants C_0 and C as follows:

$$C_0 = \max_{\mathbf{p}} |s_i(\mathbf{p}) - s_0(\mathbf{p})|$$

and

$$C = \max_{\mathbf{p}} \sum_{1 \leq i \leq n} \left(s_i(\mathbf{p}) - s_0(\mathbf{p}) + C_0 \right).$$

Since the scoring rule is non-trivial, the payments (scores) are not constant and therefore $C > 0$. Notice that $s_i(\mathbf{p}) - s_0(\mathbf{p}) + C_0 \geq 0$ and that

$$\sum_{i \leq i \leq n} \left(\frac{s_i(\mathbf{p}) - s_0(\mathbf{p}) + C_0}{C} \right) \leq 1.$$

The mechanism $M(S)$ is now defined as follows.

- The mechanism takes as input the values x_i for each of the alternatives $1 \leq i \leq n$. We assume that $x_i \geq 0$ for all i and that $\sum_{1 \leq i \leq n} x_i \leq 1$.
- Define $a_i(\mathbf{x}) = \left(s_i(\mathbf{x}) - s_0(\mathbf{x}) + C_0 \right) / C$. As observed above, $\sum_i a_i(\mathbf{x}) \leq 1$ and $a_i(\mathbf{x}) \geq 0$.
- Define $P(\mathbf{x}) = -\left(s_0(\mathbf{x}) + (1 - \sum_i x_i) C_0 \right) / C$

Proposition 2. *If S is strictly proper, then $M(S)$ is strictly truthful.*

Proof. The utility of a player playing this mechanism and reporting $\mathbf{x}$ is

$$\left(\sum_i x_i a_i(\mathbf{x}) \right) - P(\mathbf{x}).$$

This is the same as

$$\sum_i x_i \frac{\left(s_i(\mathbf{x}) - s_0(\mathbf{x}) + C_0 \right)}{C} + \frac{\left(s_0(\mathbf{x}) + \left(1 - \sum_i x_i\right) C_0 \right)}{C},$$

which is

$$\sum_i x_i \frac{\left(s_i(\mathbf{x}) + C_0 \right)}{C} + \left(1 - \sum_i x_i \right) \frac{\left(s_0(\mathbf{x}) + C_0 \right)}{C}.$$

If S is strictly proper then, it is strictly proper under the translation by C_0 and scaling by C. Thus the utility of the player in the mechanism is strictly maximized by reporting truthfully. □

A.4 Strong truthfulness

Let us define $m(x)$-strongly proper scoring rules as follows:

Definition A.1. *A scoring rule with scores $s_i(\mathbf{p})$ is $m(\mathbf{p})$-strongly proper if for every $\mathbf{p}$, $\tilde{\mathbf{p}}$*

$$u(\mathbf{p},\mathbf{p}) - u(\mathbf{p},\tilde{\mathbf{p}}) \geq \frac{1}{2}m(\tilde{\mathbf{p}})\,||\mathbf{p}-\tilde{\mathbf{p}}||^2,$$

where

$$u(\mathbf{p},\tilde{\mathbf{p}}) = \sum_i p_i s_i(\tilde{\mathbf{p}}).$$

Theorem A.2.
- *Let M be an $m(\mathbf{x})$-strongly truthful mechanism. Then $S(M)$ is an $m(x)$-strongly proper scoring rule.*
- *Let S be an $m(\mathbf{p})$-strongly proper mechanism. Then $M(S)$ is an $m(x)/C$ strongly truthful mechanism.*

Proof. The first part is immediate from the fact that utilities are precisely preserved under the transformation from mechanisms to scoring rules. For the second part, suppose that for scoring rule S

$$u^S(\mathbf{p},\mathbf{p}) - u^S(\mathbf{p},\tilde{\mathbf{p}}) \geq \frac{1}{2}m(\tilde{\mathbf{p}})\,||\mathbf{p}-\tilde{\mathbf{p}}||^2,$$

Then by the construction above

$$u^{M(S)}(\mathbf{x},\mathbf{x}) - u^{M(S)}(\mathbf{x},\tilde{\mathbf{x}}) = \frac{u^S(\mathbf{x},\mathbf{x}) - u^S(\mathbf{x},\tilde{\mathbf{x}})}{C} \geq \frac{1}{2}\frac{m(\tilde{\mathbf{x}})}{C}\,||\mathbf{x}-\tilde{\mathbf{x}}||^2.$$

□

A.5 Application to some standard scoring rules

- Logarithmic scoring rule: the translation doesn't work because C_0 is unbounded.
- Quadratic scoring rule: $s_i(\mathbf{p}) = 1 + 2p_i - ||\mathbf{p}||^2$. Then $s_1(p) - s_0(p) = 2(p-(1-p)) = 4p-2$ which is between -2 and 2. Thus $C_0 = 2$ and $C = 4$. This translates into $a(x) = x$, which has $m(x) = 1$.
- Spherical scoring rule: $s_i(\mathbf{p}) = p_i/||\mathbf{p}||$. Then $s_1(p) - s_0(p) = (2p-1)/\sqrt{p^2+(1-p)^2}$ which is between -1 and 1. Thus $C_0 = 1$ and $C = 2$. This translates into

  $$a(x) = \frac{1}{2} + \frac{2x-1}{2\sqrt{x^2+(1-x)^2}}.$$

  The $m(x)$ value is half that of the spherical rule.